

Constructing the Universal Library

Elisabeth A. Jones

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2014

Reading Committee:

Joseph Janes, Chair

Cheryl Metoyer

Philip Howard

Sarah Elwood-Faustino

Program Authorized to Offer Degree:
The Information School

© Copyright 2014

Elisabeth A. Jones

University of Washington

Abstract

Constructing the Universal Library

Elisabeth A. Jones

Chair of the Supervisory Committee:

Associate Professor Joseph W. Janes

Information School

The expanded access to millions of books provided by large-scale digitization initiatives (LSDIs) like Google Books and the Open Content Alliance (OCA) has the potential to reshape myriad social structures; yet, despite this potential, they remain underassessed as social phenomena. This dissertation seeks to provide a foundation for such assessment by situating LSDIs in historical context. Specifically, it argues that LSDIs derive from the same basic urge to provide free, widespread public information access that spurred the creation of the public library movement in nineteenth century America, and that through parallel examination of these phenomena, the older can help to illuminate the newer.

Methodologically, the dissertation comprises a nested comparative case study analysis of the early years of the two LSDIs already noted and two early American public library systems – the Boston Public Library and the Carnegie Libraries – employing a theoretical lens informed by structuration theory and social construction of technologies. The research interrogates the motivations, intended user base and collection scope, and initial implementation of each initiative, as well as the relationships between these elements. In order to investigate these questions for the two public libraries, primary archival research was carried out at three physical archives and online. For the digitization initiatives, eighteen semi-structured interviews with project leadership were conducted, and supplemented by other primary and secondary source accounts. Data analysis for all cases followed an analytical-inductive approach, in which all 197 primary source documents were qualitatively coded in two iterations.

The dissertation provides an in-depth examination of the motivations, definitions, and implementations underlying each case, with a full chapter devoted to each one. It then concludes by drawing parallels and contrasts between the four cases along each of these

lines. The themes explored in the conclusion include (1) the inequalities of power and mediation among stakeholder groups in each case; (2) the role of personal passions, principles, and pragmatics in motivating these projects; (3) the boundaries on the “universal” user and the “comprehensive” collection; and (4) the procedural and structural features that seem to characterize this type of project, especially in terms of leadership styles, standardization and systematization of procedures and infrastructures, the issue of structural persistence, and the initial absence of guidance for users.

To Rosemary & Lee Jones

*Thank you for expecting the world from me,
and for supporting me so strongly when I reached for it.*

Acknowledgements

This dissertation could not have been completed without the help and support of many individuals and institutions. Among these, I would especially like to thank:

My committee chair, Joseph W. Janes, for being my constant cheerleader, critic, and mentor.

My other committee members: Cheryl Metoyer, for keeping me honest, just like she promised; Phil Howard, for asking tough questions yet always leaving me feeling inspired; and Sarah Elwood, for maintaining a constant, positive presence throughout.

Paul Courant and the University of Michigan Libraries, for funding my last few years of writing, providing me with a wealth of new experiences and ideas, and helping me find both the beginning and the end of this massive project.

My eighteen amazing interviewees, for taking the time out of their busy schedules to be interviewed for this study, and for doing so in such a gracious, informative, and often entertaining way.

All of the wonderful archivists and librarians who guided me through my historical research process, especially Sean Casey and the other staff at the BPL Rare Books and Manuscripts Department, Jane Gorjevsky at the CCNY Archive, and John Haynes at the Library of Congress.

The Coalition for Networked Information, the Association for Research Libraries, Intel, the UW Information School, and Beta Phi Mu, for funding various pieces of this work.

Sarah Caldwell, for her many insights and hard work in helping me overhaul my coding.

My UW iPhD colleagues, and especially Parmit Chilana and Natascha Karlova, for keeping me on my toes, and for helping me stay saner and happier over these past several years.

The many iSchool faculty beyond my committee who have mentored and advised me over the years, especially Hazel Taylor, Allyson Carlyle, Adam Moore, and Mike Eisenberg.

The iSchool staff who were there to help every step of the way, especially Alpha DeLap, Wendie Phillips, Ann Corboy, and Ann Warner-Smith.

Adam Taplin, for doing a phenomenal job transcribing my interviews; and Cathy Wagner, for making a trip to the Library of Congress on my behalf.

Fred Stutzman, for inventing Anti-Social, and thus allowing me to actually get some writing done sometimes.

The baristas at Aster, Neptune, and so many others, who kept me in much-needed lattes.

And of course, Brian Danford and the rest of my family, who have been my foundation through all of this. I couldn't have done it without you.

Table of Contents

INTRODUCTION	1
1. NEED FOR THE STUDY	3
1.1 <i>Lack of Conceptual Work on LSDIs</i>	3
1.1.1 Finding the Universe	4
1.1.2 Drilling Down to Scholarly Conceptualization.....	4
1.1.3 Leaving LSDIs to Journalists and Lawyers.....	6
1.1.4 Describing Different Parts of the Elephant.....	6
1.2 <i>Lack of Contemporary Conceptual Work on American Public Libraries</i>	8
1.2.1 Mythos vs. Analysis	8
1.2.2 Local vs. Systemic.....	9
2. RATIONALE FOR THE COMPARISON	9
2.1 <i>Aspects of Interest</i>	9
2.2 <i>Boundary Cases</i>	12
2.3 <i>Contrasts</i>	15
3. RESEARCH QUESTIONS: FIRST FORMULATION	17
4. THEORETICAL FOUNDATIONS	18
4.1 <i>IDEs as Sociotechnical Systems</i>	19
4.2 <i>Power and Mediation</i>	21
4.3 <i>The Unique Attributes of Information and Information Access Technologies</i>	23
4.4 <i>The Structuration of Complex Information Systems</i>	24
4.4.1 The Duality of Structure	25
4.4.2 Power and Constraint	26
4.4.3 The Role of Information in Structuration.....	29
5. RESEARCH QUESTIONS: SECOND FORMULATION	30
RQ1 – <i>Why?</i>	32
RQ2 – <i>What?</i>	33
RQ3 – <i>How?</i>	33
METHODOLOGY	35
1. CASE SELECTION	37
1.1 <i>Large Scale Digitization Initiatives</i>	37
1.1.1 Google Books	38
1.1.2 The Open Content Alliance.....	39
1.2 <i>Early American Public Libraries</i>	40
1.2.1 The Boston Public Library	40
1.2.2 The Carnegie Library Program.....	42
2. DATA COLLECTION	44
2.1 <i>Early American Public Libraries</i>	44
2.2 <i>Large-Scale Digitization Initiatives</i>	45

3.	DATA ANALYSIS	48
3.1	<i>Phase 1: Prior to Data Collection</i>	49
3.2	<i>Phase 2: During and After Data Collection</i>	50
3.3	<i>Phase 3: Secondary Analysis</i>	52
	CASE 1: BOSTON PUBLIC LIBRARY (1848-1865)	53
1.	HISTORICAL SYNOPSIS	55
2.	MOTIVATIONS.....	61
2.1	<i>(Adult) Education</i>	61
2.2	<i>Egalitarianism & Uplift</i>	63
2.3	<i>Moral/character Development</i>	64
2.4	<i>Secondary Motivations</i>	65
2.4.1	Municipal Prestige and Competitiveness	66
2.4.2	Economic Development	67
2.4.3	Public Demand	67
3.	DEFINITIONS	69
3.1	<i>Users</i>	69
3.1.1	Intended Composition	69
3.1.2	Quantity	77
3.2	<i>Collections</i>	79
3.2.1	Size.....	79
3.2.2	Composition	80
4.	IMPLEMENTATION	84
4.1	<i>Structures</i>	85
4.1.1	Building.....	85
4.1.2	Stacks.....	90
4.1.3	Catalogues	92
4.2	<i>Processes</i>	95
4.2.1	Management.....	95
4.2.2	Acquisitions.....	98
4.2.3	Cataloging	100
4.2.4	Public Service	101
4.2.5	Protecting the Collections	105
5.	CONCLUSION	109
	CASE 2: THE CARNEGIE LIBRARY PROGRAM (~1889-1899)	111
1.	HISTORICAL SYNOPSIS	112
2.	MOTIVATIONS.....	118
2.1	<i>Carnegie's Motivations for Philanthropy</i>	118
2.1.1	Wealth as Virtue, Virtue as Duty	119
2.1.2	"Nothing for Nothing"	120
2.1.3	Evolution, not Revolution	121
2.2	<i>Carnegie's Motivations for Giving <u>Libraries</u></i>	123

2.2.1	Providing a Ladder for the Aspiring	123
2.2.2	Personal History and Affection for Books	124
2.2.3	Shifting Motivations Over Time.....	126
2.3	<i>Recipient Community Motivations</i>	127
2.3.1	Local Pride.....	127
2.3.2	Education.....	128
2.3.3	Social Palliative	129
3.	DEFINITIONS	130
3.1	<i>Users</i>	130
3.1.1	Intended Composition.....	131
3.1.2	Quantity	140
3.2	<i>Collections</i>	142
3.2.1	Size.....	143
3.2.2	Composition	144
4.	IMPLEMENTATION	146
4.1	<i>Processes</i>	147
4.1.1	Early Paternalism	147
4.1.2	Corporatization.....	148
4.2	<i>Structures</i>	157
4.2.1	Building Planning.....	158
4.2.2	Stacks Arrangement	164
5.	CONCLUSION	166
CASE 3: GOOGLE BOOKS LIBRARY PROJECT (2004-2012)		168
1.	HISTORICAL SYNOPSIS	171
2.	MOTIVATIONS.....	179
2.1	<i>Google's Motivations</i>	179
2.1.1	Mythos and Mission.....	180
2.1.2	Profit motive (or not)	183
2.1.3	Get More Data.....	186
2.2	<i>Libraries' Motivations</i>	187
2.2.1	Mission.....	188
2.2.2	Pragmatic Aims	194
3.	DEFINITIONS	200
3.1	<i>Intended Users</i>	200
3.1.1	The Global Public	202
3.1.2	Favored Users	203
3.1.3	Not-so-Favored Users	208
3.2	<i>Collections</i>	211
3.2.1	Size.....	211
3.2.2	"Comprehensiveness"	214
3.2.3	Composition	217
4.	IMPLEMENTATION	224

4.1	<i>Processes</i>	224
4.1.1	Selection and Ordering	226
4.1.2	Library Pre-processing	229
4.1.3	Transit and Scanning	231
4.1.4	Digital Delivery	237
4.1.5	Constraints on use	239
4.2	<i>Structures</i>	245
4.2.1	Viewing Books (or Parts Thereof) on Google.....	245
4.2.2	Library Hosting and Serving	250
4.2.3	Standards.....	253
4.2.4	Invisible structures.....	260
5.	CONCLUSION	263
	CASE 4: OPEN CONTENT ALLIANCE (2005-2011).....	265
1.	HISTORICAL SYNOPSIS	269
2.	MOTIVATIONS.....	275
2.1	<i>“Brewster, Brewster, Brewster”</i>	276
2.1.1	The Collector	276
2.1.2	The Tech Entrepreneur	277
2.1.3	The Open Access Visionary	279
2.2	<i>Principles</i>	280
2.2.1	Anti-Googlism	281
2.2.2	Openness	284
2.2.3	Procedural Transparency	286
2.2.4	Preservation.....	287
2.3	<i>Pragmatics</i>	288
2.3.1	Improve Technologies	289
2.3.2	Keeping Up with Technology (“Modernization”).....	291
2.3.3	Cost Reduction.....	292
3.	DEFINITIONS	292
3.1	<i>Users</i>	292
3.1.1	The Anti-Definition	293
3.1.2	The (Internet) Literate User.....	295
3.1.3	The Maker.....	296
3.1.4	The Machine.....	298
3.2	<i>Collections</i>	299
3.2.1	Size.....	299
3.2.2	“Comprehensiveness”	303
3.2.3	Composition	307
4.	IMPLEMENTATION	314
4.1	<i>Processes</i>	314
4.1.1	Administrative Structure (or lack thereof)	314
4.1.2	Scanning.....	315
4.1.3	Constraints on end-use	320

4.1.4	Transparency in Practice	322
4.2	<i>Structures</i>	324
4.2.1	Primary Interfaces: Archive.org and OpenLibrary.org	324
4.2.2	Platform Approach.....	331
4.2.3	Standards	333
5.	CONCLUSION	335
CONCLUSION		337
1.	MOTIVATIONS.....	337
1.1	<i>Positioning “Key Stakeholders” Across Cases</i>	339
1.1.1	The Boston Public Library	339
1.1.2	The Carnegie Library Program.....	341
1.1.3	The Google Books Library Project	342
1.1.4	The Open Content Alliance.....	343
1.2	<i>The Spark: Personal Passions</i>	345
1.3	<i>Grand Ideals: Motivations Based on Principle</i>	347
1.4	<i>Brass Tacks: Pragmatic Motivators</i>	349
1.4.1	Competitiveness	349
1.4.2	Financial Gain	351
2.	DEFINITIONS	352
2.1	<i>Boundaries on “Everyone”</i>	354
2.1.1	Age.....	355
2.1.2	Class	356
2.1.3	Race.....	357
2.1.4	Gender.....	358
2.1.5	Literacy/Literacies	358
2.1.6	Digital divide	359
2.2	<i>Boundaries on “Everything”</i>	361
3.	IMPLEMENTATION: PROCESSES AND STRUCTURES FOR ACCESS AT SCALE	364
3.1	<i>Leadership</i>	366
3.2	<i>Standardization and Systematization</i>	369
3.3	<i>Structural Residue</i>	373
3.4	<i>Physical vs. Intellectual Access: The Question of Guidance</i>	376
4.	LIMITATIONS.....	377
4.1	<i>Data Lacunae</i>	377
4.2	<i>Subjectivity</i>	380
4.3	<i>Comparability</i>	380
5.	LOOKING FORWARD: THE FUTURE OF THE [UNIVERSAL] LIBRARY	381
APPENDIX A: CODING FRAMEWORK.....		385
APPENDIX B: PRIMARY SOURCE DOCUMENTS.....		387
APPENDIX C: INTERVIEW PROTOCOL (GENERIC).....		398

APPENDIX D: CONTACT SUMMARY FORM	403
APPENDIX E: GOOGLE BOOKS LIBRARY PROJECT TIMELINE	405
APPENDIX F: OPEN CONTENT ALLIANCE TIMELINE	409
WORKS CITED	415

Introduction

The mythos surrounding public access to information is powerful. If only we can provide the masses with enough quality information, the story goes, we can improve democracy, the economy, and the human condition. More information, provided across all the typical barriers that divide people – class, gender, race, nationality, etc. – will equalize educational opportunities, and thereby increase equality in general. It will allow the lower classes to pull themselves up by their bootstraps; it will open windows into new realms of knowledge for those interested in exploring them; it will promote the establishment of an informed populace, capable of full participation in discourse and civil life in a democratic society.

It's an appealing picture, isn't it? A free, equal, democratic society: just add information.

This dissertation casts more empirical light on the specific sorts of institutions that have been built in the service of this utopian dream. In particular, this dissertation examines two sets of phenomena from two very different historical periods: first, the nineteenth-century American public library movement and second, the current push toward large-scale (even universal) book digitization by institutions like Google and the Internet Archive. I contend that nineteenth-century public libraries and current large-scale digitization initiatives (LSDIs) have employed much the same strategy in pursuit of the information-utopia: providing largely self-directed access to as much quality information as possible, to as many people as possible, free of charge. This study will be the first to systematically compare these two phenomena, and will specifically focus on the social motivations underlying particular exemplars of each phenomenon, and the processes through which those exemplars were defined and constructed to live up to their respective rationales.

But first, a bit of context.

The term “large-scale digitization initiative” was coined in 2006, to describe two then-recently-launched efforts to digitize all the books in the world: the Google Books Library Project and the Open Content Alliance (Lynch). The Google Books Library Project (GBLP) was announced in 2004 (under the name “Google Print”), as a two-pronged effort: on one hand, Google would work with publishers to provide full-text online searching of their books (with a limited preview available to the public for free); on the other, Google would work with major research libraries with the ultimate goal of scanning every book ever published. Books in the public domain would be available to read online or download in their entirety, free of charge; books in copyright or of uncertain rights status would be full-text searchable, but not readable online (depending on the rights and the agreements with publishers, Google might show a preview of a number of pages, of just a few snippets, or of no text from the book at all – just the basic bibliographic information) (Google 2011a, Google 2011c). In essence, this is still what Google is doing now.

The Open Content Alliance (OCA) was announced a year after the GBLP, to provide a strategic counterpoint to what many saw as Google's closed, centralized, and perhaps corporatist approach. The OCA sought to take a more transparent, collaborative, consortial approach to book scanning, led by the Internet Archive. The OCA and the Internet Archive also expressed dreams of digitizing all the books in the world – or more precisely, of offering “Universal Access to All Knowledge” (Kahle 2007b) – but took a more legally conservative and image-quality-focused approach to achieving that aim (Kahle 2005a, Kahle 2007b). In recent years, the Open Content Alliance moniker has fallen into disuse, though the Internet Archive and its partner libraries have continued the book-scanning started under its auspices, with the Internet Archive's Open Library site (<http://openlibrary.org>) little by little becoming the public face of the collaboration.

Within the last decade, these two LSDIs have already changed the information ecosystem. Between them, they have put more than 22 million books online – more than 20 million through Google Books, and more than 5 million through the OCA and its successor projects (Howard 2012, Internet Archive n.d.-b). Yet, within the academy, the only social aspect of this transformation that has received a significant degree of scrutiny is its impact on copyright law (E.A. Jones 2010): an important locus of impact, no doubt – but far from the only one. This lack of scrutiny has left us with an impoverished sense of what these projects actually *are* and what they might *mean* as transformative social phenomena.

Returning to the comparison introduced at the outset, then, the research proposed here begins to conceptualize these projects – to understand what they *are* and what they might *mean* – by situating them as cases of a broader phenomenon: the drive to democratize information access by providing substantial collections of published media for free public use. In this aspect of their aspirations – and in a surprising number of other ways – LSDIs strongly echo the precedent set in the nineteenth century by the American public library movement. In both cases, small groups of extremely wealthy individuals (mostly men) decided for varying reasons to throw their energies – and huge amounts of money – into the provision of large quantities of “quality” information to the masses, to be accessed and used in a largely self-directed, autonomous way. And underlying both phenomena has run a current of belief that by doing this, it is possible to improve society – to improve education, equality, democracy – that unproven hypothesis stated at the beginning, which some have called “the Library Faith” (Raber 1997, 67).

As Jesse Shera has noted, historical context is essential to understanding how we have arrived where we are, and to avoiding the repetition of our collective missteps and failures; just as an individual relies on memories to guide present reasoning, so a society depends on history to shed light on present challenges (1952). In the context of information democratization efforts like public libraries and LSDIs, this sort of context will shed light on numerous questions. What sociocultural factors support the initiation of such projects?

What assumptions do they reflect and promote? And to what extent have – or can – these systems achieve their universalist, even utopian aspirations?

Though this dissertation will not reach conclusions on all of these questions, it will begin to lay a foundation for thinking about them. Specifically, this research comprises a nested comparative case study analysis of four information democratization efforts – two LSDIs, two early American public libraries – in their founding years, as sociotechnical systems. This comparison will focus on the relationships between social negotiation processes and the particular structures and policies eventually offered for public use: how did these projects come to be the way they are, and why – and what impact did underlying social processes have on the structures they ultimately put into place?

1. Need for the Study

This study will fill significant gaps in the literature in at least two areas. By contextualizing and drawing conceptual connections between modern large-scale digitization initiatives and nineteenth-century American public libraries, it will contribute to understanding of both phenomena, as well as the broader social trends those phenomena reflect and project. In the sections that follow, I outline the particular areas of need that this research will fill. First, in the case of LSDIs, I will show that the literature to date – still young, and thus far dominated by journalists and legal scholars – has not yet produced coherent theoretical concepts for understanding what these projects are or what they mean. And second, I argue that much the same is true of the American public library history literature, though for different reasons – namely, (a) the library history field's tendency toward unexamined adulation of their subject, and (b) the existing literature's strong focus on narrow, internal topics, often to the exclusion of broader work on libraries' role and meaning as social and cultural institutions.

1.1 Lack of Conceptual Work on LSDIs

Despite the immense transformative potential of LSDIs, few rigorous social scientific analyses of these projects or their potential have yet emerged. Though the projects have received copious coverage in the popular press, trade literature, and legal venues, several articles over the past few years have noted the ongoing lack of rigorous scholarly analysis of LSDIs as social and cultural phenomena (Leetaru 2008, Dougherty 2010, E.A. Jones 2010, Chen 2012).

In work presented elsewhere (E.A. Jones 2010), I reviewed the existing literature on LSDIs, and based on this review, argued that this body of work did not yet present a coherent holistic concept of what LSDIs actually *are*, as sociocultural (or sociotechnical) entities, outside of the legal issues they raise. And although a few years have passed since this review was first executed, the landscape changed only slightly since, with somewhat more articles beginning to appear reflecting bibliometric and linguistic analyses of LSDI

collections (e.g., Michel, et al. 2011, Wilkin 2011, Chen 2012). As such, to situate what follows, I outline the original literature review and its findings below.

1.1.1 Finding the Universe

The literature search for this analysis was conducted using nineteen library databases¹ selected based on University of Washington Library subject specialist recommendations for eleven disciplines.² Within those databases, two very general keyword searches were executed, with some variances based on the structures of specific databases.³ An initial pass through the results eliminated duplicates, clearly irrelevant articles, letters to the editor, and book reviews. Then, what remained was compared against Charles Bailey's "Google Books Bibliography" (2010), and all articles on that list that had not already been retrieved were added. This produced a starting value of 1306 articles. Though this process may not have retrieved (and indeed, did not retrieve) a comprehensive set of writing about LSDIs, my own prior experience with this area indicates that the proportions of article types that it revealed would likely remain valid within that larger universe.

1.1.2 Drilling Down to Scholarly Conceptualization

Once a relatively inclusive set of literature had been retrieved, four successive sets of filters were applied to find the subset of the literature most likely to be helpful in developing a holistic conceptualization of LSDIs as social phenomena. The first filter had to do with article type. As shown in Figure 1, 70% of the articles found were too brief to display any serious depth (less than 3 pages), 6% were journalistic (geared toward news, description, or opinion, rather than original research), and 2% were secondary analyses (discussions of previous articles, conferences, or reports). In fact, only 22% (286 references) could be called primary analyses (original, scholarly contributions).

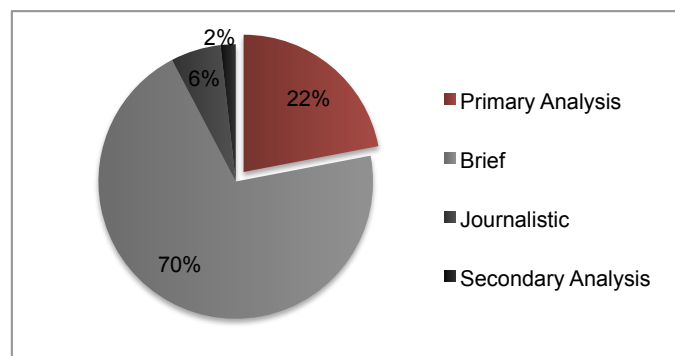


Figure 1: All Articles Retrieved, by Article Type

¹ Academic OneFile, Academic Search Complete, ACM Digital Library, America: History & Life, Business Source Complete, Communication & Mass Media Complete, Compendex, EconLit, Education Fulltext, ERIC, HeinOnline, Inspec, LISA, LLIS, LISTA, PAIS, PsycInfo, Sociological Abstracts, and Web of Knowledge.

² Subject pages used: Business, Communication, Computer Science, Economics, Education, Engineering, History of Science, Information Science, Political Science and Public Affairs, Psychology, and Sociology.

³ (1) KW: "google book*" or "google print" or "open content alliance," YR: ≥2004; (2) SU: Library materials - digitization AND KW: google (or, in databases sans subject headings: KW: library AND digitization AND google).

Working from that 286, the second filter applied was the publication type. Publications appearing in traditional serials were classed according to categories assigned by the UlrichsWeb serials index: Academic/Scholarly, Trade, and Consumer. Non-serials-based publications were classed according to broader categories: Book/Report, Conference Paper, Government Publication, and Unpublished. Then, because the goal was to gain perspective on *scholarly* conceptualization of LSDIs, these categories were used to filter out 16% of the remaining articles – those that were either unpublished or published in Trade and Consumer venues (Figure 3). Doing so left 240 articles.

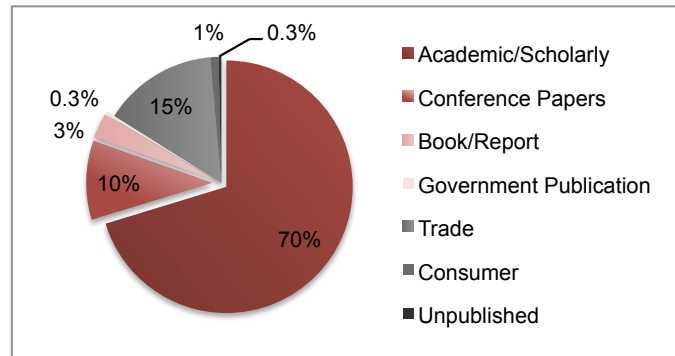


Figure 3: Original Contributions, by Publication Type

A third aspect of this body of literature that was considered was the topicality of the works: that is, were LSDIs (a) the article's primary focus, (b) used to illustrate some broader point, or (c) simply mentioned in passing, as a relevant but tangential issue? In fact, within the 240 remaining works, fewer than half were actually principally focused on the topic of LSDIs (Figure 2). Because LSDIs themselves are the primary focus here, this allows us to eliminate the combined 55% of articles that do not focus on LSDIs as phenomena, but rather as examples or illustrations of some other point. This leaves 108 articles.

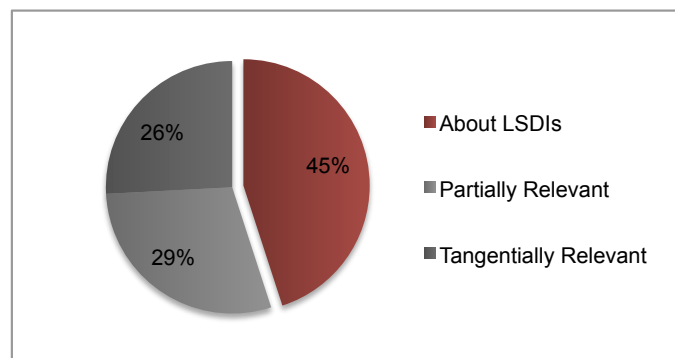


Figure 2: Original, Scholarly Contributions, by Relevance

Finally, there is the issue of the Google lawsuit. The copyright lawsuits brought by the Association of American Publishers and the Authors Guild against the Google Books Library Project, the various permutations of settlement agreements in those lawsuits, and the eventual fair use decision favoring Google have been the subject of many analyses. However, these analyses tend to be less concerned with advancing conceptual understanding of LSDIs as independent phenomena than with using

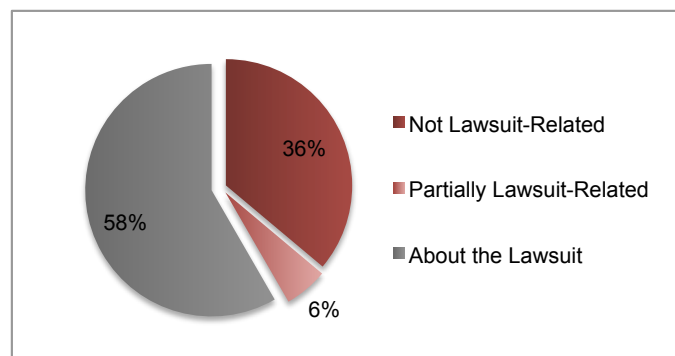


Figure 4: Original, Scholarly Contributions, About LSDIs, by Degree of Focus on Lawsuit/Settlement

the lawsuit as a lens through which to analyze copyright and/or fair use laws. Thus, I also excluded those articles focusing *exclusively* on the lawsuit – more than half of those remaining (Figure 4). That done, we are left with just 45 references out of the original 1306 that qualify as original scholarly contributions principally focused on LSDIs, not dealing exclusively with the lawsuit – a mere 3.4%.

Even preliminarily, this analysis provides two significant insights. First, on a general level, the academy (perhaps unsurprisingly) has fallen far behind the journalistic and legal communities in thinking and writing about LSDIs. And more specifically, within the tiny set of relevant, scholarly literature that does exist, little conceptual cohesiveness has thus far emerged. These insights are discussed below.

1.1.3 Leaving LSDIs to Journalists and Lawyers

Of the 387 articles found that were longer than three pages, a combined total of 196 (51%) were either journalistic or legal. And though both journalists and legal scholars have produced important and illuminating pieces on LSDIs – indeed, some of the most incisive and useful pieces of any (e.g., Roush 2005, Toobin 2007, Grimmelmann 2009a, Samuelson 2010) – neither community has shown a strong interest in systematic, theory-driven conceptualization of these projects' nature and meaning. And there are good reasons for this. Journalism, which focuses on providing the public with timely and comprehensible information, tends not to be as interested in theoretical foundations or methodological rigor as academic research. And law reviews, generally speaking, tend to privilege questions of jurisprudence and legal procedure over consideration of the broader social meaning of the phenomena that give rise to those questions. This literature should by no means be ignored by present or future scholarly analysis of LSDIs – quite the reverse – but on its own, it does not provide adequate concepts for understanding the social role of LSDIs, beyond the immediate facts of their development and their specific implications for copyright law.

1.1.4 Describing Different Parts of the Elephant

The second point that emerges from the above analysis relates to the conceptual diversity of the 45 pieces of original scholarly literature about LSDIs that remain in the end. These articles, far from painting a cohesive, unified picture of LSDIs, seem to fall into roughly seven broad (and admittedly subjective) categories:

1. Library and preservation implications (Lackie 2005, Maxymuk 2005, Milne 2005, Sandler 2005, Karle-Zenith 2006, Milne 2006, Hahn 2008, Nielsen 2008, Rieger 2008, Schneider 2008, Ceynowa 2009, Willett 2009).
2. Metadata issues (Duguid 2007, Mimno and McCallum 2007, Jackson 2008, Lu, et al. 2008, Magdy and Darwish 2008, Mann 2008, Neubert 2008, Wang, et al. 2008).

3. Holistic and/or conceptual analyses (St. Clair 2005, Coyle 2006, Jeanneney 2006, Zhao 2006, Lackie 2008, Leetaru 2008, St. Clair 2008, Dougherty 2010).
4. Descriptions or development of technological infrastructure (Choudhury, et al. 2006, Feng and Manmatha 2006, Reddy 2006, Langley and Bloomberg 2008, Schilit and Kolak 2008, Vincent 2008).
5. Analyses of LSDIs' burgeoning collections (Lavoie, et al. 2005, Lavoie and Schonfeld 2006, McShane 2008, Kousha and Thelwall 2009, Dougan 2010, E. Jones 2010).
6. Business/economic assessments (Rosenblatt 2005, van Dijk 2005, Ciriaci and Quaglione 2009).
7. Postulations on LSDIs' potential role in scholarship (Courant 2006, Martin 2008).

From these works, we can learn a great deal about LSDIs: their implications for the use and custodianship of written culture; the challenges posed by attempts to organize their massive collections, enhance their searchability, and increase their cross-platform usability; the strengths and weaknesses of the current collections for particular domain applications and scholarship more broadly; and the projects' potential impacts on the publishing industry and others involved in information dissemination. Indeed, eight of these works (category no. 3, above) even deal to varying extents with exploration of what LSDIs *are* as complex, culturally-situated sociotechnical projects.

Still, to a great extent, these subsets of the primary literature on LSDIs do not operate in the same conceptual space. Instead, they resemble the tale of the blind men and the elephant, wherein each man touches a different part of the creature – the trunk, the tail, the ear, etc. – and thus they emerge with several radically divergent ideas of what the elephant as a whole must look like.⁴ As it stands, the librarian looks at an LSDI and sees a massive digital extension of the library, with implications for preservation, document delivery, service provision, and so forth. The computer scientist, looking at the same object, sees an ideal laboratory for innovation in image standards, metadata generation, and cross-platform translation. The historian or the music theorist sees a third object still: a tool or resource with potential utility for work in her specific subject area. And on it goes, for university administrators, competing or complementary business leaders, and more.

Eliminating the diversity of these different perspectives would be both impossible and undesirable. However, in light of LSDIs' potential to influence and reshape the very nature of global information access and all the social systems that surround it, more coherent, theoretically-grounded conceptualizations are necessary – hence the need for the current research.

⁴ Several versions of this tale exist; one is available at: <http://www.jainworld.com/education/stories25.asp>

1.2 *Lack of Contemporary Conceptual Work on American Public Libraries*

Intriguingly, despite the American public library's much longer history, little more has been done to conceptualize the social nature and role of that institution. Certainly, a vast amount has been written about public libraries, including several histories of libraries in general (Johnson 1970, Harris 1995, Lerner 1998) and the American public library movement in particular (Learned 1924, Ditzion 1947, Shera 1949, Thompson 1952, Lee 1966, Garrison 1979), as well as more in-depth studies of influential libraries and library systems like the Boston Public Library (Wadlin 1911, Whitehill 1956) and the Carnegie libraries (Bobinski 1969, Martin 1993, Van Slyck 1995). However, several eminent library historians – and especially Wayne Wiegand – have noted serious analytical shortcomings in this body of literature – and especially in the large percentage of it produced prior to the mid-1970s – in particular, (1) the field's tendency to mythologize its object of study, and (2) its tendency to focus on internal, specific, or localized issues, to the point of ignoring the more systemic issues at play. Each of these issues is discussed in turn below.

1.2.1 *Mythos vs. Analysis*

With regard to the first point, Wayne Wiegand suggests that much of the literature on libraries and library history has suffered from a tendency to “celebrate, not analyze” the library: to unquestioningly accept the assumptions that libraries are bastions of democratic progress, that they strive to produce an informed citizenry capable of self-perfection, and that they perform these functions on an open, comprehensive, and egalitarian basis (2000, 5). And Wiegand is not alone in this observation (e.g., Raber 1997, Malone 2000b). To some extent, this tendency has decreased in recent decades, especially since the publication of Michael Harris's “The Role of the Public Library in American Life: A Speculative Essay” (1975a), which strongly challenged the prevailing mythologies surrounding American public libraries. In it, Harris contends:

One is to believe...that the public library movement began in a passion of liberal and humanitarian zeal; yet public libraries were generally cold, rigidly inflexible, and elitist institutions from the beginning. It is also commonly believed that the origins of the public library movement testify to the power of popular democracy in this country. Yet, historically, only a very small portion of the eligible users have ever crossed the threshold of a public library. ... These contradictions give one cause to question the public library myth (1975a, 3).

And indeed, some scholars – early on, mainly Harris himself (1976, 1986) and Dee Garrison (1973, 1976, 1979); in later years, several others, including Robert Martin (1993), Douglas Raber (1997), and Abigail Van Slyck (1991, 1995, 1996, 2001) – have taken up that charge. Their work has made great strides toward moving beyond these mythologies, and toward a more critical assessment of libraries' role in history and culture. However, the more naïve mythos of the library persists alongside this work, and continues to undermine the theoretical depth of the library history field (as also noted by Wiegand 2000, 6). The current

analysis strives to continue the trajectory of the field toward a more critical stance, grounded in social theory sufficient to the task of analyzing the complex issues of power and knowledge articulation raised by the library as a social entity.

1.2.2 Local vs. Systemic

Wiegand also identifies a second, more nuanced, gap in the library history literature. That is, Wiegand suggests that the field's "greatest shortcoming" – even above its tendency toward mythologization – lies in its tendency to ignore the broader systemic and contextual issues in library history in favor of narrower, more internal topics. As he puts it,

On the one hand, [existing work in library history] concentrates too much on the library from the inside out and focuses too much on the institution, the people who practice librarianship within that institution, and the expertise used by the people within the institution itself. On the other hand, it does not concentrate enough on the library from the outside in, nor does it focus sufficiently on people who used (or did not use) the institution, why they used (or did not use) it, and whether the expertise honed within that institution mirrored a particular "language" that was inclusive for some, exclusive for others (2000, 21).

The library history literature is rife with biographies of librarians, detailed examinations of particular institutions, and analyses of the specific processes through which libraries and librarianship function. And while many of these studies are quite valuable in and of themselves, and as subsets informing specific areas, taken in their totality they remain of limited use for thinking about the library's broader role in history and culture (Wiegand 1999, Wiegand 2000). This analysis strives to help remedy this shortcoming as well.

2. Rationale for the Comparison

Stepping back from the details of LSDIs and public libraries for a moment, a question emerges: why these phenomena, and not others? If the idea is to examine efforts to democratize access to information, have there not been many others, in many other forms? Why large-scale digitization initiatives and early public libraries?

In this section, I will answer this question in two stages: first by more precisely defining the particular features of interest in this study, and second by showing how other historical and current attempts to democratize information access fail to exhibit these features in one way or another.

2.1 Aspects of Interest

There are many lines along which comparisons might be drawn to either LSDIs or public libraries – ways of serving the public, modes of organization, considerations for building infrastructure, and so on. However, this study focuses on the similar ways in which these projects strive toward a common end – the democratization of access to information. Many efforts have been undertaken in service of this admittedly nebulous, often Enlightenment-

tinged goal – from the establishment of public education systems to the creation of mass media. Still, along this parallel, LSDIs and public libraries share a particular set of commonalities that I argue tends to set them apart as an interesting pairing for comparison. These commonalities, in fact, reveal a set of facets useful for teasing out similarities and differences between a broad range of information dissemination efforts, as I have illustrated elsewhere (Jones and Tennis 2012). In the context of this study, these facets can be laid out as a decision tree, progressively excluding myriad other types of information dissemination systems that will not be explored based on particular features of interest. In particular, I suggest that the features that set LSDIs and public libraries apart as an especially useful and interesting comparison are threefold:

1. A central goal is to provide *self-directed access to as many high-quality textual/communicative objects as possible*, in a relatively *stable* setting.
2. The target audience is explicitly *everyone*, without restrictions based on admission fees, subscriptions, users' personal characteristics, etc.
3. The information provided is, at least for the most part, *free to the user* (though it may not be free to the service provider or to other access intermediaries)

Focusing in on these three features, one can systematically exclude a diverse array of potential alternative cases for comparison, as shown in Figure 5, below. The figure, as noted, can be read as a decision tree: the narrowing criteria (e.g. whether access is self-directed or not, whether the collections are general or specific) run down the left side of the page, in diamond-shapes. Each narrowing criterion leads to two or three options: one option in each case reflects some part of the threefold description above, and leads to the next narrowing criterion; other options that do not fit the description are represented by the center column of multicolored octagons, which in turn point to some examples of the phenomena they exclude, in the rectangles along the right side. Phenomena are color-coded by the first criterion to exclude them (also connected to them by a solid line); if they could also be excluded by later criteria in the decision tree, this is represented by a dotted arrow leading from that criterion to the phenomenon. The structure of the discussion below follows the flow of the diagram, briefly exploring some of the cases that fall short of the features above in each of the listed ways, and ultimately arguing that large-scale digitization initiatives and public libraries, if not the only exemplars of every piece of the description, are at least relatively uniquely fitted to it. This discussion helps to delineate the boundaries of the current study: what is being considered here, what is not, and why?

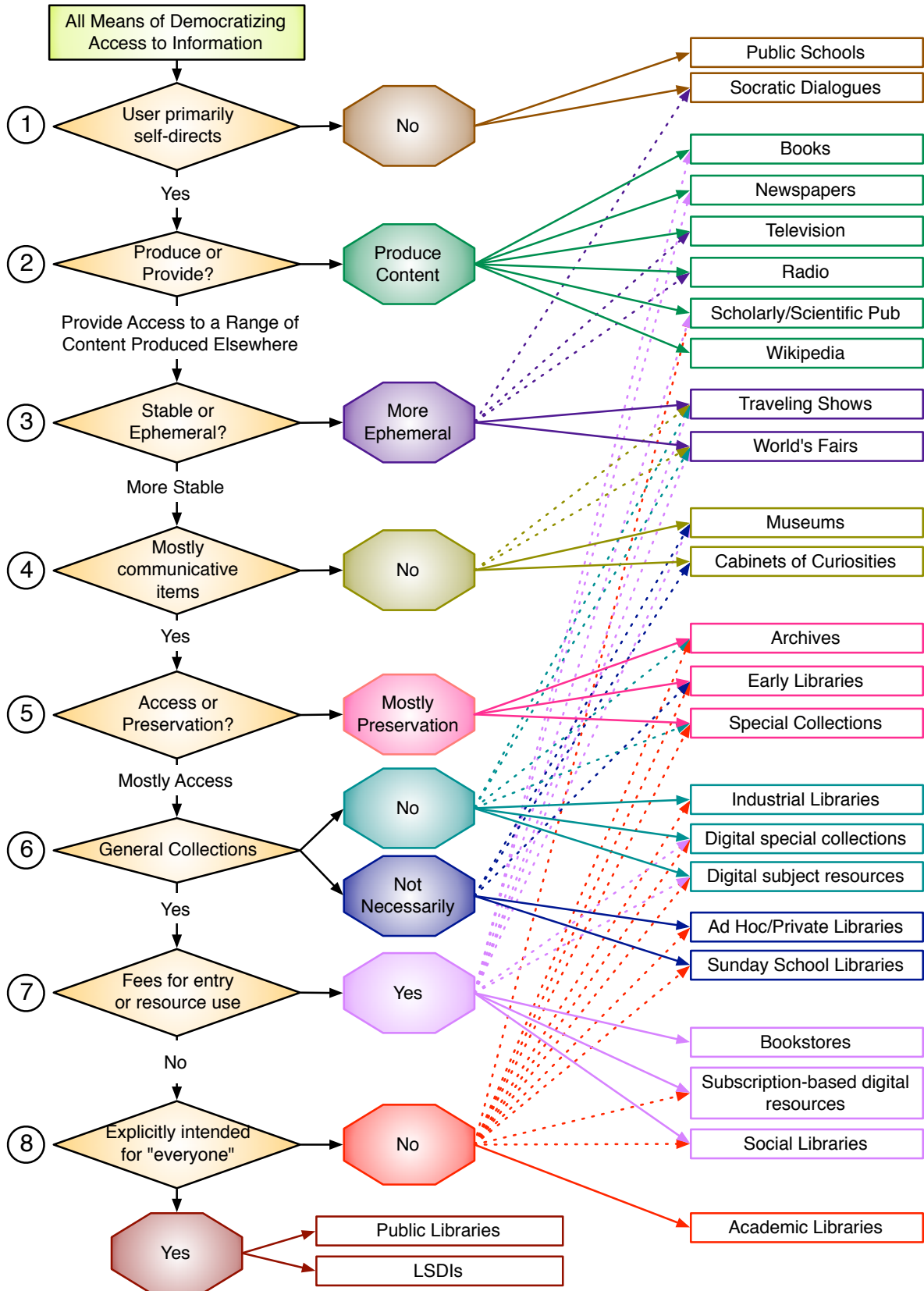


Figure 5: Boundary Analysis of Current Study

2.2 Boundary Cases

One key feature differentiating information dissemination systems is the concept of self-direction (#1 in figure). Libraries, museums, and some types of communications media (e.g. newspapers, various forms of Web content) allow the user to follow their own paths through the material; oversight or direction, though in some cases available, are not required for these phenomena to exist as such. In general, one can use a library or a museum entirely solo; though librarians or docents may be available if you need them, you are not usually *required* to make use of their services except for specific, not especially information-intensive functions, such as retrieval of items from closed stacks, or charging books out for circulation. This differentiates these phenomena from avenues of information democratization where interaction is more fundamental – most notably, public education systems, but also smaller scale phenomena such as the chautauquas of late nineteenth-century New England (Rieser 2003) or classical Socratic dialogues. Within these phenomena, some level of information-intensive dialogue and/or guidance is *intrinsic*; indeed, without such interactions, we might question whether a school or a debate would still qualify for its label. This is not to deny a certain kinship between these efforts and more self-directed forms of information access, however. Indeed, historically, support for public libraries in the nineteenth-century United States grew directly out of support for the then-new public education system – the thinking being that citizens should have public resources available to them with which to continue their education beyond their public school years (e.g., Wadlin 1911, xvii, 107, Ditzion 1947, 13-14, Shera 1949, 195-6, Lee 1971, 121). However, despite this connection, schools and libraries remain very different genres of information access.

A second differentiating aspect of information dissemination systems is the decision of whether to *produce* information or to *provide* information produced and validated elsewhere (#2 in figure). The strategy of the efforts examined here is decidedly the latter: to provide access to a range of information produced by others, rather than participating directly in information resource production. This sets these projects apart from phenomena like publishing and mass media, where the focus is on the production and dissemination (generally through sale) of resources like books, newspapers, television or radio programming, etc., as well as from newer variants of the production strategy, such as blogs or Wikipedia. Though information conduits such as television networks, newspapers, or blogs can, and often do, provide access to a wide range of information, their core purpose is not generally to assemble already-created resources, but to produce new ones.

Within the set of phenomena that do provide access to a range of content, there is a further divide between those that are more durable, or fixed and those that are more fleeting, or fluid (#3 in figure). As Levy has suggested, fixity and fluidity exist as a spectrum for information resources: different sorts of information exhibit differing levels of fluidity, and

within that fluidity exhibit varying rates of change (1994). Consider the various ways we experience music, for example. Some options, like buying CDs or MP3s, exhibit a good deal of fixity from the user's perspective – once one owns a song, one can generally replay it as much as one likes, and it will be the same every time. Other options, however, are much more fluid and fleeting: for example, streaming audio, and its antecedent, radio, fly past and disappear into the ether unless captured – if such capture is even feasible given contemporaneous technologies. These more fleeting conduits, which present their information synchronously, then let it disappear, also fall outside the scope of this study. This characteristic also serves to exclude phenomena such as World's Fairs and traveling shows, which were extremely valuable in transferring a broad array of knowledge across sometimes vast distances, but which generally disappeared without a trace (aside from the occasional building) once their run was completed (Hobhouse 1937, Harrison, et al. 1980, Davis 2002, Ganz 2008). Further, this criterion provides an additional exclusion factor for forms of media such as television and radio, whose content generally disappears as it plays, unless purposefully captured.

This leaves us, essentially, with stable institutions or projects geared toward providing access to diverse collections of materials – largely libraries, museums, archives, and their digital equivalents. However, this set can be pared down according to the definition as well (#4 in figure). Museums (and their more *ad hoc* cousins, cabinets of curiosities) tend to focus on non-textual/communicative objects: their focus is on artifacts, rather than documents. Indeed, even where museums do contain documents, they tend to be valued and used more as historical or cultural objects than as information resources (e.g., Martin 2007, 84) – and few museums take texts (even broadly defined) as the core focus of their collections. Yet, as will be discussed in greater detail in the explication of the theoretical framework, documents have a special kind of cultural valence because of their ability to record and reproduce human speech (Levy 2001). It is thus collections of documents that will be of particular interest here.

Still, even within collections of documents, there is segmentation. One important line along which this segmentation occurs is the preservation/access divide (#5 in figure). Most libraries founded prior to the advent of mechanical printing, as well as most contemporary archives and special collections, have focused much more strongly on preservation than on access. The written materials they house tend to be rare (or even unique), valuable, and difficult to duplicate, and more access means a greater chance that these precious, often irreplaceable artifacts will be damaged or lost (e.g., Posner 1940, Johnson 1970, Lerner 1998). These collections thus also fall outside the scope of interest here: they are not so much embodiments of information democratization as they are of information conservation – an equally noble aim, and one critical to the perpetuation of access into the future, but not the focus of this research.

A second line of segmentation within collections of documents relates to the contents of those collections (#6 in figure): many collections, including most early mechanics', mercantile, private, and Sunday School libraries, as well as most extant "digital libraries" or "digital collections," had or have a quite specific subject focus. For mechanics' and mercantile libraries, this was linked to the particular industry in which they were situated; for example, a library located in a cotton mill would likely provide mainly pragmatic information relevant to textile production, though the collection might extend slightly into other areas (Shera 1949, 229-35). And more *ad hoc* book collections, like Sunday School libraries and liberally-opened private book collections like the one Andrew Carnegie famously used in his youth (e.g., Carnegie 1889a, 689, Van Slyck 1995, 9, Jones 1997, 5), were just that – *ad hoc*. Dependent on the particular needs of the institution in which they were embedded and/or on the tastes of the book purchaser, these collections made no pretense of universality (Shera 1949, Van Slyck 1995). Collection specificity in digital libraries has somewhat more complex origins, but basically, it boils down to cost. Digitization, especially in its early years, has been expensive; thus, only relatively circumscribed collections have been selected to be digitized. The early history of digital libraries is thus to a great extent the history of subject-specific projects like *Making of America*,⁵ *Early English Books Online*,⁶ and *MEDLINE*,⁷ as well as smaller, collection-specific projects like the digitization of Northwestern University's Winterton Collection of East African Photographs,⁸ or UC-Berkeley's online exhibit on the 1906 San Francisco Earthquake and Fire⁹ (e.g., Greenstein and Thorin 2002). Since the focus here is on general – even purportedly or aspirationally "universal" – collections, these more specialized types of libraries and other projects can be set aside as well.

The final winnowing factors, which both eliminate many of the remaining types of collections and provide secondary rationales for eliminating many of the phenomena already excluded, is the extent of access, as delineated in the second and third major parts of the definition. That is, the phenomena of interest here are set apart by being explicitly *intended* for the use of *all* (#8 in figure), and by offering up the use of at least a large proportion of their collections *free of charge* (#7 in figure). Most of the boundary cases already discussed fail to meet one or both of these criteria: most books and newspapers cost money; World's Fairs, traveling shows, and many museums charge(d) admission; expensive subscriptions have been required for most digital libraries; and most libraries and archives (physical or digital) have been designed for the use of a specific user base, rather than the general public. To these, we can then add social and subscription libraries

⁵ <http://quod.lib.umich.edu/m/moagrp/>

⁶ <http://eebo.chadwyck.com/marketing/about.htm>

⁷ <http://www.ncbi.nlm.nih.gov/pubmed>

⁸ <http://repository.library.northwestern.edu/winterton/>

⁹ <http://bancroft.berkeley.edu/collections/earthquakeandfire/index2.html>

like the Library Company of Philadelphia, which required the payment of membership dues in order to use their collections (Shera 1949, 32-33); bookstores, where the ultimate point is to sell books, and not to provide unlimited free use; and even academic libraries, where collections have been selected and policies designed to meet the particular needs of a specific (and generally geographically local) scholarly community, rather than those of the public at large.¹⁰

This leaves us with the two phenomena suggested at the outset – public libraries and large-scale digitization initiatives. Of the projects that have been undertaken to this point, they alone exemplify every part of the description above: both attempt to provide everyone within their orbit with free, self-directed access to an extremely diverse array of information resources. And in an interesting way, they also share a focus on providing *quality* information: public libraries through the selection of the “better books” to compete with the “mischievous, poor reading [then] indulged in” (“Report of the Trustees of the Public Library of the City of Boston” 1852, 18); LSDIs through the selection of books *in general* – the medium containing “the highest-quality knowledge” to supplement (or perhaps compete with) the variably reliable information available on the open Web (Toobin 2007, quoting Sergey Brin). For the sake of brevity, I will employ the term “Information Democratization Efforts” or “IDEs” as an umbrella to refer to these two phenomena alone. Though, as explained above, many other phenomena share certain attributes with these projects – and though many of these could also quite reasonably be described as information democratization efforts – for the purposes of this research, the term “IDE” will only extend to describing early American public libraries and large-scale digitization initiatives, as delineated above.

2.3 Contrasts

Early American public libraries and modern large-scale digitization initiatives share important characteristics in terms of motivation and strategy: that much should be clear from the discussion above. However, it would be remiss not to note that there do exist several important differences between these two types of information universalization scheme. Three of these stand out as especially striking, and potentially informative for the current research: (1) the projects’ respective subsidiary motivations, (2) their models for ongoing funding, and (3) the legal issues they have faced. I will briefly discuss each of these in turn before moving on to the research questions.

Though LSDIs and early American public libraries share an important *central* goal – to

¹⁰ Many academic libraries are open to the public – indeed, for academic libraries participating in the U.S. Federal Depository Library Program, this is required by law – however, serving the public is not the primary mission of an academic library. Rather, they exist to serve the needs of the academic institution in which they reside; to support the research and instruction being done there.

broaden access to information – the *secondary* factors motivating their creation diverge greatly. For public libraries, unlike LSDIs, the desire to extend the public education system, paternalistic efforts to elevate public morals, and a perceived need for public vocational training all played significant roles (Shera 1949). The first two of these in particular emerge as significant in both the Boston and Carnegie cases described in later chapters here. Further, public libraries were (and are) strongly connected to the communities in which – and *for* which – they were built; the needs and priorities of the community influenced everything from the design of the building to the job description of the librarian to the library’s book-purchasing strategies. The situatedness of public libraries within their communities forms a key difference between these institutions and LSDIs, which operate under the assumption of a more global audience. And LSDIs also exhibit secondary motivations not present in public libraries: most notably, one LSDI, Google Books, is at least partially intended to serve Google’s underlying profit motive, and this motivation runs alongside the company’s hopes to support universal access.¹¹ This corollary motivation leads Google to engage in various activities that generally are not pursued by libraries, such as facilitating book sales alongside lending, and also gathering and retaining massive amounts of tracking data on the readers it serves.

Google’s profit motive, moreover, points to a second contrast between LSDIs and public libraries: their respective models for ongoing funding. From the outset, public libraries were explicitly intended to be public, tax-supported entities (Shera 1949). Though they initially relied heavily on private donations, the core goal was to create, as it were, libraries ‘by the people, for the people.’ This differs strikingly from Google Books, which has been funded mostly by Google itself all along, with certain costs related to book scanning borne by the company’s library and publisher partners, and others offset by book sales. Moving forward, it seems likely that sustaining the library digital copies and collections will become incorporated into the budgets of the libraries and consortia involved, while Google’s collections will remain Google-funded (at least, until the company loses interest in supporting it). And although both the OCA and its successor initiatives at the Internet Archive seem committed to remaining nonprofit, neither has indicated any apparent plans to move from donation to tax support; rather, the IA appears to subsist mainly on grants and donations, and this does not seem likely to change (Open Content Alliance 2009, Internet Archive n.d.-a). Though both public libraries and LSDIs relied heavily on the support of wealthy benefactors to get them started, public libraries quickly turned to public funding, while LSDIs have employed a greater diversity of funding models – and the long-term viability of this constellation of support remains uncertain.

¹¹ This contrast has become especially clear with the release of Google’s “ebookstore” (<http://books.google.com/ebooks>).

Finally, public libraries never had to face the kind of opposition that LSDIs have on copyright grounds (see, e.g., Samuelson 2010). In order to provide a book to many people, American public libraries simply had to purchase one copy (or, often, several copies) of that book; at that point, the first sale doctrine in U.S. copyright law allowed it to be lent to others.¹² In order for an LSDI to provide a digital book to many users, however, requires at least two additional – and potentially problematic – steps: first, a digital copy of the book must be created through scanning, and second, every time a user requests the book, a further copy of that digital file must be made in the user’s computer memory (Litman 2001). This difference means that, depending on the book’s copyright status, legal restrictions may apply at many more points in LSDIs’ operations than public libraries’. This, in turn, has caused legal headaches for LSDIs that paper-based libraries never had to face – especially for Google, which continues to face the ongoing aftereffects of a class-action lawsuit brought against it by the Association of American Publishers (AAP) and the Authors Guild (AG) in 2005, for making scans of paper books without asking permission first.¹³ And even in the case of the OCA, whose legal approach was at least initially much more conservative, copyright remained a major issue: the labor required to ensure that everything scanned is either in the public domain or explicitly permitted by the rights-holder would be simply unnecessary to build a paper-based collection of similar materials for public use.

3. Research Questions: First Formulation

In the discussion to this point, I have striven to demonstrate that (a) LSDIs are under-conceptualized as social phenomena, and (b) LSDIs and early American public libraries share a constellation of characteristics that make a comparison of these two phenomena particularly intriguing with regard to thinking about emergent forms of large-scale public information access as social phenomena. This dissertation will not attempt to fully remedy problem (a): thorough conceptualization of the structures and social implications of LSDIs (as for public libraries before them) would be a task for many dissertations, if not many careers. Instead, the current research will focus on three elements of this conceptualization: first, the specific social motivations that have spurred the initial creation of individual IDEs; second, the internal definitions of individual IDEs espoused by key stakeholders, especially in terms of intended users and collection scope; and third, the processes through which

¹² This, of course, is changing in the current e-book environment, in which the model of public library book lending has begun to encounter strong and vocal opposition (as well as technologically-enforced limitations, via DRM) from the publishing community.

¹³ Much could be said – and has been said – about the lawsuit, and I won’t repeat it all here. The basic status as of this writing is that a settlement between Google and the original plaintiffs was proposed, revised, and ultimately rejected by a judge, mainly for going too far beyond redress of the plaintiffs’ grievance into making new law – which, he noted, is the province of Congress (*Rejected Settlement Opinion* 2011). The AAP privately settled their part of the lawsuit with Google in late 2012 (Sporkin 2012b); the AG lawsuit continues under appeal, having initially concluded with a summary-judgment finding of fair use (on the part of Google) the following year (Chin 2013).

individual IDEs have sought to manifest their respective internal definitions and motivations in terms of structures and policies.

These areas of interest are informed by a theoretical framework drawn primarily from Giddens's theory of structuration (1984), but also informed by work in sociotechnical systems, and communication and media studies (e.g., Innis 1951, Pinch and Bijker 1984, Hughes 1987, Holmes 2005). This framework is detailed in the next section; I introduce the questions here briefly and plainly, in order to anchor that discussion:

RQ1 ("Why"): In the four cases under consideration, what were the initial motivations for pursuing each project expressed by and among key stakeholders (those with some capacity to influence what happened)?

RQ2 ("What"): How did those responsible for casting the initial shape of each project conceive of what they were doing – or hoping to do? Put differently, what was each project's native concept of itself?

RQ3 ("How"): How did each project initially go about achieving the goals and definitions outlined in RQ's 1 and 2, particularly in their implementations of particular structures and processes?

These questions, as noted above, are in no way exhaustive, and this formulation of them lacks certain specifying elements drawn from the theoretical framework. However, stating them in a rough form here will help to tether the discussion of theoretical foundations below to the basic outlines of what is being studied.

4. Theoretical Foundations

Large-scale public information access projects like those examined here implicate a wide range of broad, yet complex issues. Technology, history, and communication; democracy, equality, and citizenship; business, academia, and "the public" – all these play a role. In order to effectively analyze such projects, one must employ a theoretical framework capable of accounting for not only their philosophical vastness, but also their empirical intricacy. In particular, I suggest that such a framework must exhibit three core attributes:

1. *Bridging the social-technical (and functionalist-interpretative) divide:* The framework must avoid assessing IDEs solely as technological/functional systems on the one hand, or solely as human/social systems on the other. Instead, it must account for the ways in which these two aspects of IDEs both simultaneously manifest and intermix, and avoid privileging either above the other.
2. *Accounting for power dynamics:* The framework must incorporate a nuanced account of power relations, capable of illuminating the role of phenomena like mediation,

translation, and negotiation, which recur in and contribute to shaping information universalization schemes.

3. *Accounting for the unique properties of information resources:* The framework must account for the important ways in which information systems and resources differ from other sorts of systems and resources present in social life. Information resources (or “documents”), as Levy has suggested, are “talking things:” unlike, say, a toaster or a spoon, information resources have been endowed with the capacity to repeatedly impart messages to other human beings; in essence, to *speak* (2003, 36). This sets them apart from other artifacts in terms of their social role, and thus, situating such resources and the systems we build around them requires an especially nuanced theory of sociotechnical structures and dynamics.

Based on these requirements, I propose that Giddens’s theory of structuration, augmented in certain respects to tailor it more closely to the subject matter, will provide a robust framework to guide the analysis of large-scale information democratization efforts. In the sections that follow, I will first describe the relevance and importance of each of the three criteria above in greater detail, and then discuss how structuration theory either meets these criteria or might be extended to do so.

4.1 IDEs as Sociotechnical Systems

During the early years of the public library movement, library leadership focused principally on the *social* aims and ideals they hoped the library would serve – among them, a more educated populace, an increase in equality of opportunity, and stabilization of the socioeconomic status quo. Concern with architecture – not only of buildings but of classification systems and institutional hierarchies – while extremely important to the attainment of these social goals, did not gain equal attention until much later (“Report of the Trustees of the Public Library of the City of Boston” 1852, Dana 1897, Wadlin 1911, Ditzion 1947, Garrison 1979, Van Slyck 1991). In digital libraries, the reverse was true: designers initially focused almost exclusively on the technical aspects of information provision – how to make books accessible as bytes – with little thought given to the social assumptions and implications embedded in those systems (Levy 2000, 25). Both of these historical trajectories illustrate why it is important not to mark off the social from the technical aspects of IDEs: since such efforts inevitably create *sociotechnical* systems – systems where social assumptions and interpretations become embedded in the structural properties of the system, and vice-versa – researchers interested in understanding them must account for both of these aspects simultaneously, or risk missing important parts of the picture.

In recent years, the library literature has increasingly recognized the utility of conceptualizing digital libraries (DLs) in particular as sociotechnical systems (STS). Julian Kilker and Geri Gay, for example, observe that “[not] only are DLs intricate *technological*

systems comprising scanning, storage, transmission, display, and printing components, they are also embedded in complex *social systems*, comprising librarians, engineers, funders, scholars, and general users" (1998, 60, emphasis added). Van House, Bishop, and Battenfield echo this perspective, noting that digital libraries are "composed of people, activity, artifacts, and technology," and thus require "an analytical stance that 'privileges neither the social nor the technological and in which neither is reducible to the other'" (2, citing Levy 2003). And though these authors limit their claims to *digital* libraries, they are equally true of physical/traditional libraries. Physical libraries, after all, are not only *socially-embedded* institutions populated by librarians, funders, scholars, and general users; they are also *technological* systems comprising storage, organization, access, and distribution components. Further, the claims also remain true when applied beyond digital *libraries* to other systems providing access to digitized works, such as the Internet Archive Text Archive or Google Books. Thus, whether assessing a very recent IDE like Google's or a deeply historical one like the Boston Public Library, an account of the project's dual social-technical nature is not only relevant, but necessary.

This requirement significantly narrows the range of theoretical frameworks capable of dealing with IDEs. Most of all, it necessitates the rejection of any framework that elevates the functionalist over the interpretative, or vice-versa. On one side, as O'Day and Nardi note, functionalist thinking tends to conceive of technological systems centrally as *tools*, and thus leads mainly to assessments of "how people, tasks, and technologies fit together," without situating those elements in broader social, organizational, or political contexts (O'Day and Nardi 2003, 66-7). And furthermore, functionalism tends to imply a broadly teleological theory of social dynamics, wherein structural elements – including technological systems – become deterministic with regard to human action. But this will not do in a framework for IDEs. In fact, the structural features of a sociotechnical system are only deterministic to the extent that agents engaged with that system *take actions that make them so*: the library only educates if its patrons use it to inform themselves; the digitization initiative only broadens access to information if users strive to grasp the information that it conveys. As several STS theorists have noted (notably Latour 1991, Akrich 1992), structures may provide a script, but agents determine the extent to which that script will be followed.

Still, the opposite epistemological extreme, the interpretative view of social systems, is equally flawed as a framework for dealing with IDEs. In strict interpretative thinking, the only reality is that which individuals create through their actions and utterances; knowledge is inherently subjective, and the natural world is marked off as something inherently mysterious, about which humans, given the limitations of our senses, cannot ever know the "truth" (Giddens 1984, 1-2). Yet such interpretative theories are, as Giddens eloquently suggests, "founded...upon an [untenable] imperialism of the subject:" in this model, the purposive actions of individuals explain all social life, while the structural

configurations and constraints within which those individuals move are largely ignored (1984, 2). Such neglect of the influence of structure and structural constraints in the context of information access systems would produce an almost laughably incomplete analysis. Technology, after all, may not have a *purely deterministic* influence on the social dynamics that surround it, but it certainly has *some* influence. Indeed, sociotechnical systems theorists have described a range of impacts that technology can have in various settings: it can “define a framework of action” for users (Akrich 1992, 208); extend, shape, rework, or reproduce existing social and economic arrangements (Bijker and Law 1992a, 11); provide rules and resources for individuals to draw upon in interpersonal interaction (DeSanctis and Poole 1994, 125); or many other possibilities.

Indeed, contrary to the conclusions that would be drawn from either purely functionalist or purely interpretative perspectives, human action and technological structure each play a non-negligible role in shaping the reality of sociotechnical systems, as technology and action recursively and iteratively influence one another. Neither action nor structure is reducible to the other, and both must be equally well represented in an adequate theoretical framework for assessing IDEs as sociotechnical systems.

4.2 *Power and Mediation*

The second element necessary in forming an adequate theoretical account of IDEs is a nuanced concept of power relations. Such an account is required for two reasons. First, without an account of power, one cannot accurately deal with the variation in actors’ abilities to influence a system’s design and evolution. Within IDEs, as in any complex sociotechnical system, different groups of relevant actors have different capacities to influence the shape and direction of that system, based on their position in surrounding power structures as well as their level of access to the system’s decision-making and design processes. An account of power is fundamental to explaining the sources and dynamics of such inequalities. And second, one must account for power because it is a fundamental motor of social life: it is not only an instrument wielded when decisions are made and conflicts resolved, but is a pervasive characteristic of *all* social relationships and structures. It is, as Giddens suggests, “a routine element of the instantiation of conduct in social reproduction” (1984, 15-6) – and assessing IDEs without accounting for it would produce a seriously impoverished analysis.

In fact, it is just such an inadequate account of power that I would argue renders Pinch and Bijker’s popular framework for assessing sociotechnical systems – “Social Construction of Technologies,” or “SCOT” (introduced in Pinch and Bijker 1984) – insufficient as an independent basis for analysis of IDEs. In many ways, SCOT offers an appealing perspective for such analysis. For one thing, SCOT beautifully captures the recursive relationship between technologies and social actors described above. In contrast to earlier,

more teleological analyses, Pinch and Bijker describe the process of innovation as “an alternation of variation and selection,” resulting in “a ‘multi-directional’ model,” wherein technologies take shape through a dialectical process among designers, potential users, existing technical capacities, and other factors (Pinch and Bijker 1984, 411). Additionally, SCOT is one of very few theoretical frameworks that simultaneously incorporates this recursion *and* deals explicitly with the social construction of *technologies*, as opposed to more abstract scientific theories and knowledge structures (as in, for example, Berger and Luckmann 1966, Sismondo 1993, Maines 2000). On these points, at least, SCOT provides quite a nuanced perspective on the dynamics of sociotechnical systems – certainly preferable to a strictly functionalist or interpretative approach.

However, as noted, the framework loses much of its potential explanatory capacity by failing to offer even the most basic account of power. SCOT views technological innovation as a process of negotiation among “relevant social groups” – groups that Pinch and Bijker assume will be “quite straightforward” to identify, and will all have equal influence upon the innovation process (1984, 414). These assumptions, however, significantly oversimplify the power dynamics of the situation: in any decision-making process, technological or otherwise, some groups or individuals will hold more sway than others, or will have more or less access to decision-making fora; indeed, even relatively uniform groups of individuals are rarely fully uniform in opinion (Kilker and Gay 1998, 67-68, Klein and Kleinman 2002, 30-31, Shin 2006, 93). As Klein and Kleinman assert,

Implicitly, SCOT assumes that groups are equal and that all relevant social groups are present in the design process. This fails to adequately attend to power asymmetry between groups. Some groups may be effectively prevented from participating in the design process at all.... Some groups may not be groups at all but may be a diverse collection of subgroups for whom some actor claims to speak.... The process of artifact development may be rife with intergroup...and intragroup conflict (2002, 30-31).

In the context of IDEs, these issues are readily apparent. Consider, for example, the early public library movement: not only did the wealthy, male, WASP leadership hold more sway over design than any other group, but they argued heatedly amongst themselves, disagreeing on, among other things, the proper sorts of books to provide, the degree to which those books should be made accessible to the public, and the centrality of the educational motive to the library’s mission (Ditzion 1947, 17). To cast these individuals monolithically as a single “relevant social group,” and then to set them up as equal in power to, say, factory workers – who often felt unwelcome in and rejected by the libraries (Garrison 1979, 39-42) – would be not only factually incorrect, but analytically misleading.

Further, by ignoring the role of power, a rigid application of SCOT theory would blind the analyst to other important social dynamics involved in the development of complex technologies – especially, in the context of IDEs, the role of mediation. Kilker and Gay’s

evaluation of the *Making of America* project (MOA) – the sole extant study to have employed the SCOT framework to analyze an information access initiative – brings this weakness clearly into focus. Kilker and Gay note that different groups “relevant” to MOA’s design stand at varying levels of remove from the actual design process. For example, “[users], mediated by both evaluators and developers, are two levels of influence from the technology,” while librarians, able to communicate directly with the developers, stand at only one level of remove (Kilker and Gay 1998, 67). Such multiple input sources certainly add more diversity of opinion to the design process than would no “outsider” input at all; however, the ways in which these inputs *are mediated* tend to mitigate and otherwise alter the mediated groups’ ability to influence design: aside from creating a more muted voice, Kilker and Gay note, mediation can “introduce error, as in the children’s game of telephone” (1998, 67). The social processes that lead to the creation of complex public-facing initiatives like IDEs necessarily implicate a broad range of social groups, with an equally broad range of levels of remove from design and decision-making processes. An account of mediation, as a corollary to an account of power, is thus strongly indicated as an element in an adequate theoretical framework for analysis of these systems.

4.3 *The Unique Attributes of Information and Information Access Technologies*

Finally, a suitable theoretical framework for work in this area must contend with the unique attributes of information and information access technologies. Though all technologies that enter into common use have some impact on their social contexts, the impact of information technologies has the potential to be especially profound, as it exists on two separate, significant levels. Not only do information access systems, like other technologies, influence social arrangements through the provision of some practical function – like an oven heating food, or a bicycle providing transportation – they also affect societies through the messages they carry, conveyed in human language, across space and time. As Levy proposes, “documents are talking things. They are bits of the material world (clay, stone, animal skin, plant fiber, sand) that we’ve imbued with the ability to speak” (2003, 30). Documents, conceived in this relatively broad way – the information resources on which IDEs are based – are thus more than just functional objects. They have been endowed with a kind of *life*; an ability to speak human language, and convey human messages, in direct and explicit ways. For this reason, a theoretical framework that lumps information technologies in with other sorts of technologies, ignoring their unique properties, loses the ability to capture perhaps the most socially influential aspect of information access systems, relative to technologies like the oven or the bicycle: their ability to convey and shape the recipient’s interpretation of human communication.

The recognition of these unique characteristics of information resources serves less as a criterion for rejecting particular theoretical systems than as an argument for incorporating others. Levy notes that the “ability to keep talk fixed” in various forms has become “an

essential cornerstone of human social organization,” and that such communicative fixity is fundamental to all major social institutions, from science to religion to government (2003, 38). This being the case, the incorporation of theoretical work from fields dealing explicitly with the transmission of human thought through various media – especially communication studies, but also education and information science – will help to enrich understanding of the social dynamics at play in IDE design and use. For example, the idea that communication technologies are fundamentally more influential than non-communicative technologies with regard to social organization and individual perception forms a core assumption underlying the work of ‘medium’ theorists such as Innis, McLuhan, and Meyrowitz (Carey 1967, 7, Holmes 2005, 38-42). In explicating and comparing Innis’s and McLuhan’s theoretical work, Carey describes their distinction between communication technologies and other technologies in a way that strongly presages Levy:

Most instruments are attempts to extend man’s physical capacity, a capacity shared with other animals. Communications technology, on the other hand, is an extension of thought, of consciousness, of man’s unique perceptual capacities. Thus communication media, broadly used to include all modes of symbolic representation, are literally extensions of mind (1967, 7).

And further, these “extensions of mind,” in their role as the raw materials of pedagogy, become the instruments of socialization. In education, representations of past experience – that is, information resources – are placed in context and given meaning as guides for future experience: as Dewey suggests, education “is that reconstruction or reorganization of experience which adds to the meaning of experience, and which increases ability to direct the course of subsequent experience” (1916, 76). This process, at least in most literate societies, fundamentally relies upon information resources as the foundation for such reconstruction.

Social systems and the individuals who populate them are shaped and altered by the thoughts communicated within and among them. The transmission of information stands at the core of social life; in the form of education especially (whether formal or informal), it has the capacity to establish and reinscribe particular forms of social life across geographic and temporal distances. By incorporating theoretical concepts from disciplines that deal explicitly with these factors, a framework for studying IDEs will become better equipped to contend with the unique range of social influences these projects entail.

4.4 The Structuration of Complex Information Systems

With all these factors in mind, I propose that Giddens’s theory of structuration, augmented in certain respects to account for specifically technological contexts and the properties of information resources, will form a solid theoretical foundation upon which to construct a detailed analysis of information universalization schemes as sociotechnical systems. In

what remains of this section, I will describe how structuration theory meshes – or can be extended to mesh – with each of the three requirements in turn.

4.4.1 The Duality of Structure

Structuration theory centers on the reconceptualization of several *dualisms* – subject vs. object, functionalism vs. interpretativism, and the individual vs. society – as *dualities* (Giddens 1984, 25, 162). For example, rather than imagining the actions of individuals as subordinate to overarching social structures, or vice-versa, Giddens claims that “the structural properties of social systems are both medium and outcome of the practices that they recursively organize” (1984, 25) – and thus, that individual actions both *reflect* and *constitute* the social structures within which they operate. Action and structure are not separate, but form two sides of the same integrated whole. This duality helps structuration to escape the limitations of social theories that disconnect agent from structure and reify one over the other. Instead, structuration theory suggests a recursive, mutually reinforcing relationship between individuals and social systems:

Human social activities, like some self-reproducing items in nature, are recursive. That is to say, they are not brought into being by social actors but continually recreated by them via the very means by which they express themselves *as* actors. In and through their activities agents reproduce the conditions that make these activities possible (Giddens 1984, 2).

Not only do agents’ actions (re)produce the structural conditions under which they live; those structures go on to define the range of possible actions available to those agents going forward.

Social life in structuration theory thus becomes a cyclical dialogue between action and structure; between intention and consequence. Individuals go through their lives acting on particular interests that reflect their understanding of the structural and social contexts that surround them, and those actions, in turn, alter their context – often not in the ways their perpetrators intended – and those alterations influence future agents’ understandings and interests, thus creating a causal loop (Giddens 1984, 185). And so it is with IDEs. A group of individuals (or more likely, several groups) sets forth with a particular set of intentions, based on their knowledge of the world – perhaps a faith in democracy, or a desire to organize information, or a legal requirement to make money for shareholders. Out of those intentions, then, come actions: the construction of a library, the formation of a consortium, the drafting of particular contracts and agreements. Those actions leave structural residue – the library’s architecture, the inclusion or exclusion of particular partners, the policies outlined for selection and use of materials – which in turn imposes rules and creates resources that inform the possibilities for future action. Agency and structure become two parts of the same dynamic whole, engaged in a continuous dialogue that reproduces, alters, and extends the defining features of social life.

This is very much like the understanding of the sociotechnical cycle proffered by many STS theorists, and drawing upon their work will help to anchor the more general theory of structuration to the context at hand. For example, Hughes suggests that technologies are both “socially constructed and society shaping:” the process of developing new technologies is fundamentally dependent on the assumptions and experiences of “system builders” (as well as the resources they have available to them), and the physical and social environment for future thought, interaction, and innovation is in turn reshaped by the introduction of each successive technological innovation (1987, 51-53). This broadly systemic view of technological development and uptake meshes well with and aptly extends Giddens’s theory to treat sociotechnical systems in a more nuanced and specific way.

4.4.2 Power and Constraint

Power is a pervasive and fundamental element of structuration theory; indeed, Giddens flatly asserts that “There is no more elemental concept than that of power” (1984, 283). Power, in structuration theory, is the means through which all action is accomplished; the medium of social and systemic reproduction; the vehicle for domination, freedom, and everything in between (Giddens 1984, 257). In practice, Giddens suggests, individuals exercise power by drawing on “rules and resources” provided by the social structures that surround them (1984, 25). Social structures – including sociotechnical structures – thus become sources of both constraint and sanction; they limit the range of ways in which power can be exercised, yet also define the very elements one might draw upon to do so. A brief detour into Giddens’s concepts of rules and resources will help to illuminate this point.

Rules, for Giddens, can be broken down into four types: (1) habit or routine, (2) constitutive rules, (3) regulative rules, and (4) formulae. The first type, habit, is a fairly weak set of rules, with no actual sanction – as in, “as a rule, I get up at 7 am.” Nobody requires compliance with such rules, and there is not necessarily any reason for them; they are merely what happens as a matter of course (Giddens 1984, 19). The second and third types, constitutive and regulative rules, are related. Constitutive rules *create* the activity that they regulate (as, for example, the rules of a chess game create the game of chess), while regulative rules specify the ways in which activities must be carried on (as in, “it is a rule that workers must clock in at 9 am”). Giddens acknowledges that these may be two *aspects* of rules, rather than two variant *types*, but maintains the distinction to highlight the difference between rules’ role in creating meaning and their close connection with sanctions (1984, 19-20). The final type, formulae, are for Giddens the most germane to analysis of social life, because they describe generalizable procedures. Just as $a^2 + b^2 = c^2$ always defines the relationship between the sides of a right triangle, regardless of differences in the specific dimensions, social formulae apply “over a range of contexts and occasions,” and allow for the enactment

of regularized participation in social activity in particular contexts. A social formula, Giddens explains, will not “specify all the situations an actor might meet with, nor could it do so; rather, it provides for the generalized capacity to respond to and influence an indeterminate range of social circumstances” (1984, 20-22). Rules can thus appear in many forms, from formalized laws to tacitly understood social norms. And yet, one should not confuse *formalization* with *regulative capacity*. Often the most profoundly influential rules are those that are informally and tacitly held; those that actors may not even realize they are following. For example, the tacitly observed gentility norms of the early public library had both regulative and constitutive force: they both dictated behavioral requirements for library use and helped to define the nature of the institution itself (Garrison 1979). And while LSDIs may not enforce particular behavioral norms, their structures – their user interfaces, search and sorting algorithms, etc. – nevertheless have their own kind of regulative force: for example, the list of results retrieved for a particular query will present only one of many possible orderings of the relevant (or not so relevant) information – and this ordering, in turn, will affect how the user comes to understand the answer (or range of answers) to her original question.

Interestingly, this understanding of rules dovetails fairly well with an alternative formulation proffered by cyberlaw scholar Lawrence Lessig, in his discussion of the forces that regulate individual behavior. For Lessig, these four forces – Norms, Laws, Architecture, and the Market – are simultaneously distinct and interdependent:

Each can support or oppose the others. Technologies can undermine norms and laws; they can also support them. Some constraints make others possible; others make some impossible. Constraints work together, though they function differently and the effect of each is distinct. Norms constrain through the stigma that a community imposes; markets constrain through the price that they exact; architectures constrain through the physical burdens they impose; and law constrains through the punishment it threatens (Lessig 2006, 124).

Lessig, like Giddens, suggests that these regulative forces serve to both constrain and constitute action: the norms of a community simultaneously help to define what that community is (constitutive rules, in Giddens’s terms), serve as guides to appropriate behavior there (regulative rules), and suggest modes of interaction (e.g. manners, customs) that structure members’ responses to various situations (formulae). Similar points could be made about law, though it is more formalized, and architecture, though it is less visible in everyday interaction (indeed, such invisibility is one indication of a successfully functioning infrastructure (Bowker and Star 1999)). Market forces do not function in precisely the same way; yet, given the mixed market position of projects like Google Book Search – simultaneously in the business of giving things away for free and, eventually, selling those same or other very similar things – Lessig’s reminder to account for the

regulative power of market forces is well taken, and provides a useful extension to the theory.

Perhaps even more interesting for an analysis of IDEs, however, is Giddens's account of *resources*. Resources fall into two categories: *allocative* and *authoritative*, where

Allocative resources refer to capabilities – or, more accurately, to forms of transformative capacity – generating command over objects, goods or material phenomena. Authoritative resources refer to types of transformative capacity generating command over persons or actors (Giddens 1984, 33).

Allocative resources, then, include the “Material features of the environment (raw materials, material power sources),” the “Means of material production [or] reproduction (instruments of production, technology)” and “Produced goods (artifacts created by the interaction of 1 and 2),” while authoritative resources include the “Organization of social time-space (temporal-spatial constitution of paths and regions),” the “Production/reproduction of the body (organization and relation of human beings in mutual association)” and “Organization of life chances (constitution of chances of self-development and expression)” (Giddens 1984, 258). The perpetuation of any social system across time and space requires the mobilization of some combination of these resources, and such resources are also implicated in any situation of domination. Within IDEs, this conceptualization of resources becomes theoretically relevant in three ways. First, the relative distribution of allocative and authoritative resources among IDE stakeholders helps to shed light on the differences in influence wielded by the groups involved in their design. Some groups have more allocative resources than others (Google can fund and technologically support mass digitization at a level that most libraries never could) and some have more authoritative resources than others – wealthy white men in nineteenth century America wielded much more political power than other socioeconomic or ethnic groups of the time – and often these will be the same groups. Resources thus contribute to the relative “relevance” of social groups (Pinch and Bijker 1984, Klein and Kleinman 2002) implicated in IDE design. Second, on the side of impacts, IDEs explicitly seek to alter the distribution of intellectual resources within social systems: by broadening access to information, they also (ideally) broaden access to control over users’ own life chances and their own self-production and reproduction – and, ultimately, their access to material resources and the means of production. This redistribution, in turn, would exemplify Giddens’s claim that power is the medium of both domination and freedom: by appropriating the resources that effect one’s domination, one acquires the capacity to alter or even reverse the impacts of those resources upon one’s social role or opportunities. Finally, the third way that the concept of resources arises in IDEs relates to the issue of mediation: to the extent that certain groups mediate other groups’ input into technological design, they possess a particular sort of authoritative resource. Though mediators may not

have actual *command* over those for whom they mediate, they do have control over how much of the other groups' messages get through, and how accurately those messages are conveyed.

4.4.3 The Role of Information in Structuration

Finally, Giddens explicitly acknowledges the unique role that information resources and modes of information dissemination play in the transmission and reinscription of social structures. Specifically, Giddens claims that evolving modes of information transmission

[expand] time-space distancing by an 'alienation' of communication in circumstances of co-presence. Any written text becomes distanced from its author; printing is for the most part a quantitative extension of such distancing. Electronic media separate presence in time from presence in space, a phenomenon of decisive significance for contemporary forms of collectivity (1984, 203).

Just as Levy claims that the defining characteristic of documents is their "ability to hold talk fixed, so it can be repeated again and again at different points in space and time" (2003, 30), Giddens maintains that information resources, because they can be held at a distance from their authors both geographically and temporally, serve to distribute individual presence over time and space – a phenomenon which both reinscribes existing social structures and, as media types evolve, extends those structures' potential boundaries.

Further, Giddens notes that the power of information lies not only in its storage, but in its retrievability – both physical and intellectual:

Storage presumes *media* of information representation, modes of information *retrieval* or recall and, as with all power resources, modes of its dissemination. . . . All [information media] depend for their retrieval upon the recall capacities of the human memory but also upon skills of interpretation that may be possessed by only a minority within any given population (Giddens 1984, 261).

This is a key point for IDEs: information dissemination is incomplete if those at the receiving end do not have the skills to interpret what they obtain; technological access is decidedly not coterminous with intellectual access. And if only a select few recipients have the skills to obtain intellectual access to an IDE's contents, those few will gain in power relative to the many who do not; after all, it is only they who will have truly gained an allocative resource from their encounter with the IDE, and it is thus only they who will have the capacity to translate and transmit as much or as little of it as they choose.¹⁴ And because of this, mediation and power become all the more strongly implicated as necessary features of a theoretical framework for studying IDEs.

¹⁴ Indeed, this mode of using information as an allocative resource wielded by the few over the many has the capacity to shape politics and history: consider, for example, the vast power the Mayan kings wielded on the basis of their exclusive knowledge of astronomy (e.g., Rice 2004, xvii-xviii)

Information democratization efforts implicate massive – and massively complex – social phenomena, from innovation to democracy. By accounting for the duality of structure, the intrinsic importance of power, and the unique attributes of information, structuration theory goes a considerable distance toward providing an adequate theoretical framework for dealing with these complex projects and their impacts. Supplemented by the more topical theoretical concepts provided by work in SCOT, as well as certain elements of information and media theory, it becomes stronger still, and provides powerful guidance for the further specification of the research questions introduced earlier.

Still, a good theoretical framework need not account for every last detail of the phenomenon under study; indeed, it would lose all generalizability if it did. The framework need only provide, as Giddens suggests, “sensitizing devices” for thinking about and drawing conclusions from research results (1984, 326-7). So long as these sensitizing devices do not leave blind spots in areas crucial to the phenomenon under study – like those dealt with above – they will help in both framing appropriate questions for and drawing informed conclusions from the work.¹⁵ IDEs have literally millions of moving parts – technologies, people, information resources, ideologies – and perhaps no theory could completely encompass all the issues they raise. However, I believe that structuration theory, especially with the additions noted, will provide a useful device for beginning to understand both the breadth and the depth of these projects.

5. Research Questions: Second Formulation

At this point, it will be helpful to return for a moment to the research questions posed earlier, and more fully explain how these questions emerge from and relate to the theoretical framework being employed in this study, and in particular, the cycle of sociotechnical influence and the duality of structure outlined above. In this cycle, on a basic level, social processes – negotiations between individuals, assumptions based on existing social structures and individual biases, interpersonal and macrosocial power dynamics – have a significant effect on the particular shape that a given technology will take; and once constructed, that technology cycles back to influence the social settings in which it is implemented or adopted – and the shape of that influence is, at least in part, determined by the shape of the technology, which itself is socially formed. This cycle could be (fairly simplistically) visualized like so:

¹⁵ Or, as Manuel Castells advised a graduate symposium I attended in 2010, “Use theory. Don’t let theory use you.”

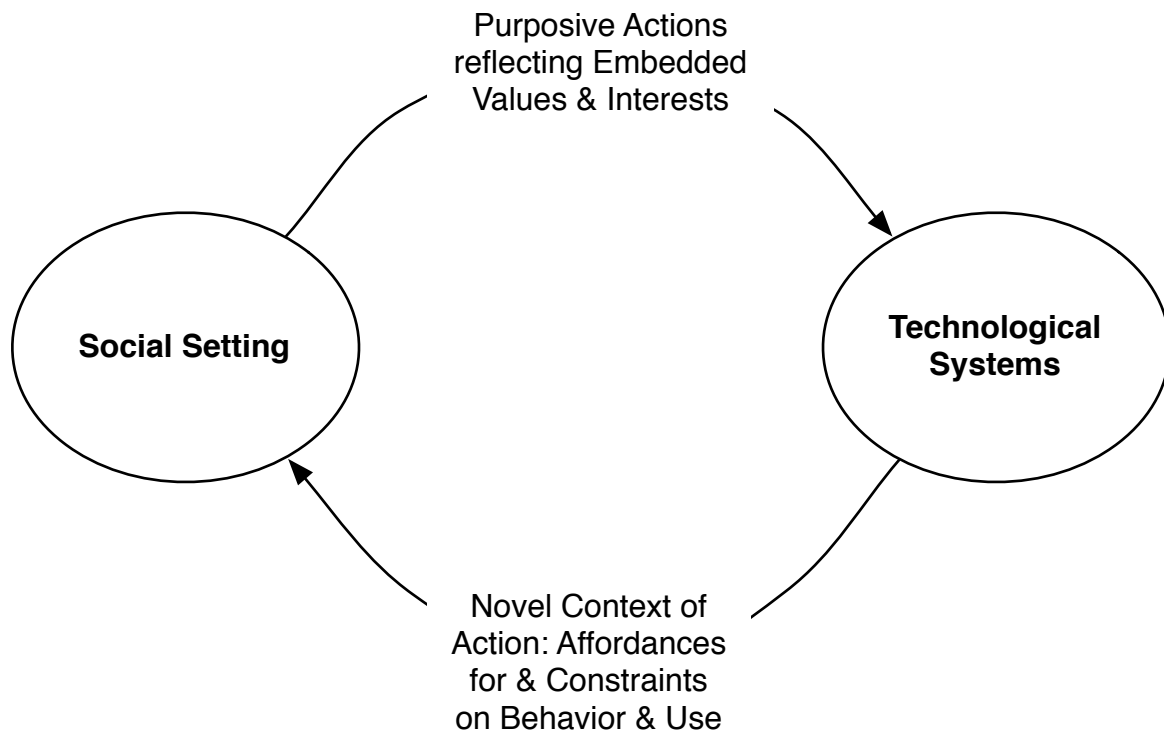


Figure 6: Sociotechnical Cycle

In an ideal world, it would be preferable to examine this cycle as an integrated whole, holding nothing fixed and accounting for all its dynamism at once (as suggested, for example, by Latour (2005)). However, particularly in the context of a doctoral dissertation, such comprehensive integration would produce far too expansive a set of issues to realistically investigate. Thus, I situate my research mainly within the upper portion of the diagram: the social setting, the purposive actions taken (which reflect embedded values and interests), and in a more limited way, the technological systems that shaped by these actions. More specifically, it will examine (1) the social motivations that have spurred the initial creation of individual IDEs (social setting); (2) the internal definitions of individual IDEs espoused by key stakeholders, especially in terms of intended users and collection scope (values and interests transposed into concrete action); and (3) the manifestations of these respective internal definitions and motivations in terms of structures and policies (systems).

The questions were stated earlier as follows:

RQ1 (“Why”): In the four cases under consideration, what were the initial motivations for pursuing each project expressed by and among key stakeholders (those with some capacity to influence what happened)?

RQ2 (“What”): How did those responsible for casting the initial shape of each project conceive of what they were doing – or hoping to do? Put differently, what was each project’s native concept of itself?

RQ3 (“How”): How did each project initially go about achieving the goals and definitions outlined in RQ2, particularly in their implementations of particular structures and processes?

At this point, each of these questions can be further specified in more theoretically-grounded terms.

RQ1 – Why?

In the four large-scale IDEs under consideration, what were the initial motivations for pursuing each project expressed by and among key stakeholders (those with some capacity to influence what happened)?

Structuration theory suggests that agents engage in a constant reflexive monitoring of the actions and structures that surround them (including their own actions) that, given a certain basic level of competence, produces an agent knowledgeable about their social setting, and broadly capable of rationalizing their actions (Giddens 1984, 5-6). Distinct from this capacity for rationalization, however, lie motivations, which Giddens describes as “potential for action rather than...the mode in which action is chronically carried on by the agent. ... For the most part motives supply overall plans or programmes...within which a range of conduct is enacted” (1984, 6). And although, as he notes, much day-to-day conduct is not directly motivated, an account of motivations is an essential element of any analysis of “strategic conduct:” that is, any analysis where “the focus is placed upon modes in which actors draw upon structural properties in the constitution of social relations” rather than attempting to explicate the chronic reproduction of structural properties without as much focus on particular agents (what Giddens calls “institutional analysis”) (1984, 288-89). In the current study, as noted above, the focus is precisely on such strategic conduct: this question exposes the motivations behind the activities of particular actors – individuals with some capacity to shape a particular IDE – as reflective of and embedded in particular social settings and value systems. Admittedly, to some extent, motivations become entangled with rationalizations in this study, particularly for the cases with interviews; this would likely happen in any situation where individuals are asked to explain their motivations after the motivated action has already occurred. Still, to the extent that the data also includes statements made at various earlier points in time, this issue has been at least partially mitigated.

Furthermore, beyond the theoretical rationale for the question, on a practical level, an understanding of the motivations of various actors within these projects serves as an illuminating backdrop against which to view the steps taken by those actors in bringing the

projects to fruition. Though these agents' purposive actions may have led in other directions than intended – as, indeed, Giddens claims is inevitable (e.g., 1984, 294) – teasing out the motivations behind them helps to describe at least the beginning of their trajectories.

RQ2 – What?

How did those responsible for casting the initial shape of each project conceive of what they were doing – or hoping to do? Put differently, what was each project's native concept of itself?

In particular:

- **How many and what kinds of people did the project's founders think would be using their creations?**
- **How comprehensive was the collection intended to be, and how was it shaped by policies regarding what to include and/or exclude?**

The theoretical rationale for this question is quite similar to that for RQ1. That is, it helps to describe the beginnings of the multiple action-trajectories that ultimately collide to produce the IDEs under study.

RQ3 – How?

How did each project initially go about achieving the goals and definitions outlined in RQ2, particularly in their implementations of particular structures and processes?

In particular:

- **What design moves were made to orient the structures (both interfaces and backend) toward their intended user base and level of comprehensiveness?**
- **What procedures were put in place to help achieve the goals outlined in RQ1 and definitions outlined in RQ2?**

A key element of the theoretical framework is the mutually constitutive relationship between actions and structures. The first two questions explore the ideation (motivations and definitions) behind actions taken toward creating novel social institutions; the third takes a closer look at the structures – both technical and administrative – put in place in order to bring those initial ideations into reality. These technical and administrative elements both reflect and will reflect back upon their social contexts, and not always (or even most of the time) in the ways that their initiators intended. In this way, as Giddens suggests,

a given set of activities...is interpreted as purposeful action. In other words, those activities are shown to be carried on in an intentional way, for certain reasons, within conditions of bounded knowledgeability. Specification of those bounds allows the analyst to show how unintended consequences of the activities in question derive from what the agents did intentionally. The interpretation involves an attribution of rationality and motivation to the agents concerned. The actors have reasons for what they do, and what they do has certain specifiable consequences which they do not intend (1984, 294).

Though the current study does not extend as far as an examination of consequences, mapping out the particular configurations of concrete actions taken in these cases will help to lay the groundwork for future work on the social effects – intended and unintended – of these massive projects for the societies in which they are adopted.



All of these questions also provide avenues to explore the power dynamics and mediation processes at play in the construction of IDEs. And as just noted, they will also lay the foundation for exploring a third set of questions – those examining the actual social *impacts* of IDEs – which, though extremely important, lie beyond the scope of the present analysis. Some questions in this area, for example, might include: in what ways have the structural properties of longstanding IDEs (public libraries) influenced the social contexts into which they have been inserted? What social groups or characteristics have particular IDEs tended to privilege or favor, and which have they excluded or discouraged? How might these historical phenomena inform a projection of the ways in which newer IDEs (LSDIs) will influence the social contexts in which they will be used, and how might this projection inform the design of these systems? These questions implicate a much larger set of issues, and would require a much more complex set of research processes to fully investigate; they may thus form a fruitful set of pathways for future research in this domain.

Methodology

In order to address the research questions posed above, this dissertation employs a nested comparative case study approach, examining the origin moments of two American public library systems and two large-scale digitization initiatives. The comparative case study method was chosen primarily based on its ability address complex questions of process and trajectory like those put forward here. As Yin suggests, case studies can be especially illuminating for research where (1) the focus is on “how” and “why” questions, (2) the researcher has little opportunity to exert control over events, and (3) contemporary phenomena are involved (2003, 5-9). Dubé and Paré build upon these contentions, noting that case studies are most useful “when a phenomenon is broad and complex, when a holistic, in-depth investigation is needed, and when a phenomenon cannot be studied outside the context in which it occurs” (Dubé and Paré 2003, 598). All of these are characteristics of the current research. Additionally, the case study has advantages over traditional historical research in terms of data triangulation and analytical depth: as Yin notes, “the case study’s unique strength is its ability to deal with a full variety of evidence—documents, artifacts, interviews, and observations—beyond what might be available in a conventional historical study” (2003, 8). By drawing in a wide variety of materials, case studies allow the researcher to triangulate different types of data in order to increase the validity of conclusions drawn about each case, as well as generalizations among them.

Furthermore, there is strong precedent for the use of case study methodologies in research which, like this study, assumes a sociotechnical theoretical perspective. Much, if not most, extant work in sociotechnical systems has utilized case studies, historical methodologies, or both to construct holistic descriptions of the social foundations and impacts of specific technical systems (e.g, Bijker, et al. 1987, LaFollette and Stine 1991, Bijker and Law 1992b, Fox 1996). And although most of this work has involved unpacking the social foundations of non-information-intensive technological systems, such as light bulbs (Bijker 1992), door keys (Latour 1991), or movie projectors (Carlson 1992), it has more recently begun to extend to more information-rich technologies such as management information systems (Orlikowski 1992, DeSanctis and Poole 1994, Orlikowski and Barley 2001) and digital libraries (Kilker and Gay 1998, Bishop, et al. 2003). The precedent for this methodological choice within research sharing a similar theoretical backing – and an increasingly similar area of focus – provides further support for the case study as a strategy for this dissertation.

Though the case study method is not as universal in research that takes structuration as its theoretical model, it is nonetheless a well-respected approach to empirical work in that area. The application of structuration theory to information technologies has been most prevalent in the management fields, especially organization studies (e.g, Orlikowski 1992, Heracleous and Barrett 2001). And as Pozzebon and Pinsonneault (2005, 1361-62) suggest,

based on their exhaustive study of the use of structuration theory in organization studies, structuration theory seems to call for a “process approach” – that is, an approach that examines sequences of events in action, rather than testing hypotheses or assessing variables against one another.¹⁶ Following on this assertion, Pozzebon and Pinsonneault also note that in the existing organization studies literature, the case study is one of the preferred strategies for data collection and analysis (alongside ethnography, grounded theory field studies, and action research) (2005, 1362). They note that these approaches share largely interpretative epistemological assumptions, a grounded, data-driven logic, and elements of narrative – all elements well within the scope of a case study approach (2005, 1362-64).

Still, the term “case study” only goes so far in articulating a strategy for research. Methodologically speaking, the term describes less a unified set of techniques than a more general approach to research. Case studies focus on understanding their given cases as integrated wholes; this is done by piecing together the interrelationships between the elements that make up each case under consideration (Schutt 2009). However, what exactly those elements *are*, and how they are discovered, may differ greatly between different case studies. The precise case selection, data collection, and data analysis methods for this study will be discussed in the sections that follow. At this point, I will simply note that the study’s overarching structure – a nested comparative case study, examining two sets of two cases – was selected for two reasons. First, and most importantly, by utilizing a *comparative* case study approach rather than focusing in on a *single* case, this research has the capacity to produce conclusions that extend beyond the specific cases it examines. As Yin notes,

Analytic conclusions independently arising from two [or more] cases, as with two [or more] experiments, will be more powerful than those coming from a single case (or single experiment) alone. [Also,] the contexts of the two cases are likely to differ to some extent. If under these varied circumstances you still can arrive at common conclusions from both cases, they will have immeasurably expanded the external generalizability of your findings, again compared to those from a single case alone (2003, 53).

Furthermore, contrasts between cases that are in other ways alike can also be informative, by highlighting the influence of the cases’ differing contexts, settings, and relevant actors. And, more generally, such contrasts can also reinforce the non-inevitability of the form taken by any single instantiation of a given phenomena: large-scale book digitization need not look like Google Books, any more than public libraries must all resemble Boston’s first

¹⁶ “Variance approaches,” using survey or experimental procedures that cast relationships in terms of dependent and independent variables (Pozzebon and Pinsonneault 2005, 1360-61), are common in applications of Adaptive Structuration Theory (DeSanctis and Poole 1994). However, as Pozzebon and Pinsonneault note, these applications tend to bear little relationship to Giddens’s original formulation of structuration theory (1361).

effort; examining two cases of each type side by side will help to illuminate the divergent possibilities.

The second reason for choosing this research design builds upon this last point. The universe of large-scale book digitization initiatives is still quite small: there have been no more than a handful of book digitization efforts of any substantial size, and there have been still fewer that have striven for anything approaching universal coverage. As a result, it can be difficult to keep in mind that not all book digitization efforts with universal ambitions need look like the largest and most famous of these efforts to date – i.e. Google Books – and that, in fact, not even that initiative need retain its current form forever. Including the Open Content Alliance thus offers a window into an alternative process for striving toward very similar goals – decentralized, nonprofit, in some ways more transparent, and not as legally tenuous – as a way of beginning to understand the issues that are and are not fundamental to the design and implementation of this type of project. The decision to include two public library systems rather than one arises partially from a similar logic, but also partially from a desire for parallelism between the contemporary and historical cases. The similarities and contrasts that this research will find between the structuration processes of the Boston Public Library and the Carnegie Libraries will help to illuminate the extent to which each of those systems was unique versus the extent to which each was reflective of the broader social or historical phenomena it represents.

1. Case Selection

The cases for this study were selected primarily based on their perceived historical significance as embodiments of the information democratizing phenomena they represent: (1) large-scale digitization initiatives (LSDIs) and (2) early American public libraries, respectively. In each case, the dissertation focuses on the *initial* form taken by the initiative: for the LSDIs, roughly their first five years, and for the public libraries, through the earliest bounded phase of their development (for the BPL, up to and slightly beyond the establishment of the first dedicated library building; for the Carnegie Libraries, up to the beginning of Carnegie's period of "wholesale" library giving (Bobinski 1969, 13)). That said, the specific rationales differ between the digitization projects and the library cases. This section will describe those rationales.

1.1 Large Scale Digitization Initiatives

The category of LSDIs is already a small one. In fact, based on the criteria outlined in the discussion of boundaries above, I would contend that only five initiatives that existed at the outset of this research in 2008 could possibly have served as cases for this study – Project Gutenberg, the American Memory Project, the Million Book/Universal Library Project,

Google Books, and the Open Content Alliance – and that the last two of these rise above the others as especially significant and interesting cases, for slightly different reasons.¹⁷

1.1.1 Google Books

Google Books would stand out as an important case purely on the basis of scale, even if it had no other interesting features (which it does – more on that next). The number of volumes it has scanned, and the pace at which has scanned them, are on a completely different scale from any of the other LSDIs listed above. The oldest LSDI, Project Gutenberg, currently includes approximately 100,000 volumes, re-typed by volunteers one by one since 1971 (Project Gutenberg 2008, Project Gutenberg 2013). The Library of Congress’s American Memory

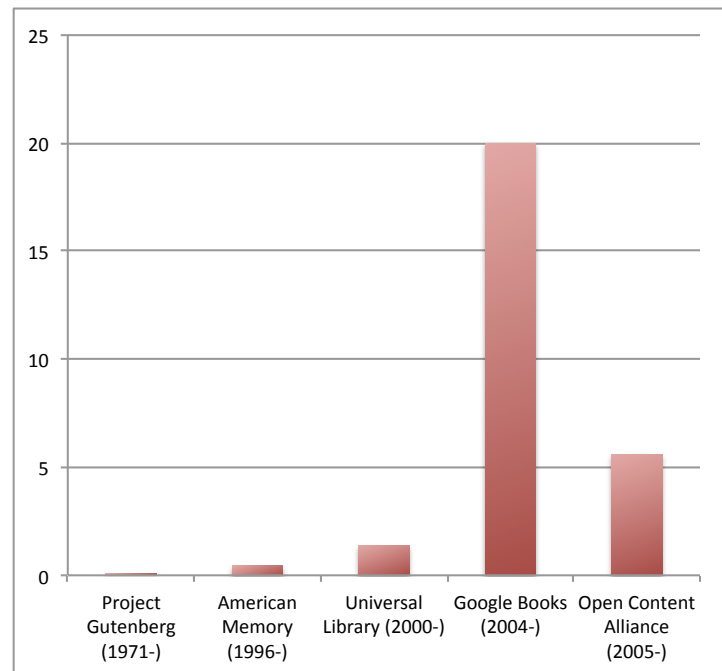


Figure 7: Volumes Scanned (in millions) as of December 2013

Project has scanned 9 million items since 1996, including a wide variety of non-book materials (e.g. photographs, sound recordings, moving images, pamphlets, maps); Gloriana St. Clair estimates that these 9 million items roughly equate to 450,000 volumes scanned (2008, 152). The Million Book/Universal Library Project scanned 1.4 million volumes between 2000 and 2008 (St. Clair 2008); though the project seems to be ongoing at least nominally, more recent data is not clearly available (as of late 2013, the project website¹⁸ did not appear to have been updated since 2008). The Open Content Alliance has scanned or integrated a relatively impressive 5.6 million volumes in just 8 years (Internet Archive n.d.-b). However, Google Books has dwarfed even this figure: in the last nine years, Google has independently scanned more than 20 million books (Howard 2012). This places Google’s average digitization rate at roughly 2.2 million volumes *per year* – nearly half as many books *every year* as the OCA has incorporated in its entire eight-year lifespan. Additionally, it is worth noting, as at least one librarian interviewed for this study did, that many of the books

¹⁷ HathiTrust (2008-present), which as of March 2012 boasted over 10 million volumes (<http://www.hathitrust.org/>), might also be included in this company. However, it is not shown here primarily because of its role as a secondary repository: the volumes it contains represent a patchwork of resources digitized through other projects – mainly Google Books – and contributed to Hathi as a secondary, nonprofit repository; it is thus in some sense the second generation of both Google Books and the OCA (as several interviewees noted).

¹⁸ <http://tera-3.ul.cs.cmu.edu/index.html>

included in the OCA corpus were actually scanned by Google and later uploaded to the OCA by individuals from the Pirate Bay (P8; more on this issue in Chapter 6). Figure 7 illustrates the vastness of the difference in scale between Google Books and all four other digitization projects, even if the OCA is given credit for the Google scans it hosts.

The scale and pace of Google Books would alone make it an interesting case. However, there are at least three other points of interest that further support its inclusion here. First, partly as a result of its enormous scale, Google Books has become by far the most well-known digitization initiative among members of the general public (e.g., St. Clair 2008) – and as such likely also the most widely used. Second, its success at getting books digitized and searchable demonstrates the viability, at least in the short term, of the private, centralized model of digitization it represents. And finally, the controversy over its copyright policy had the effect of placing Google’s Books team and library partners in a unique position to influence not just the shape of digitization, but the shape of digital copyright law for years to come. These factors, combined with the forceful influence of Google in the world in general at present (e.g., Levy 2011, Vaidhyanathan 2011), indicates that what Google is doing now will have a strong impact on both future digitization strategies and the ecosystem within which they will develop – whether because later initiatives choose to emulate or differentiate from Google, or because Google Books squelches all competition and becomes the only game in town. In either case, it is important to understand the reasoning behind Google’s strategy, and how that reasoning has shaped the information resources now being offered there.

1.1.2 The Open Content Alliance

The Open Content Alliance, though considerably smaller, makes sense as a comparative case running parallel to Google Books for a few key reasons. First and foremost, the Open Content Alliance (OCA) was founded fairly explicitly as a counterpoint to Google Books (Albanese 2007d, Leetaru 2008). Its main coordinating organization, the Internet Archive (IA), had been scanning books for years before the announcement of the OCA (Kahle 2005a); however, the announcement of Google Books seemed to light a fire under IA founder Brewster Kahle to build an alternative effort to digitize all the world’s books – one that would proceed on a more open, collaborative model than Google’s. The features that set the OCA apart from GBS form the second key reason for its inclusion as a case. The OCA differed from Google Books in at least three ways of analytical interest to this study. First, the OCA was much more decentralized, both mechanically and administratively: it relied heavily on digitization executed independently or semi-autonomously by its partners, and did not claim centralized control over the resources it gathered together (though some of its partners – notably Microsoft – did claim some control over their contributions, at least for a time) (Leetaru 2008). Though the OCA did have scanning centers of its own, it also incorporated items digitized elsewhere; for example by the

Universal Library Project (100,000+ items integrated), Project Gutenberg (36,000+), and, as previously noted, Google Books (1 million+) (Kahle 2007b, Internet Archive n.d.-b). Second, the OCA focused much more strongly on scanning each and every volume at a very high quality than Google has, and the IA continues to provide that high-resolution version directly to users – a choice which increases the detail and precision of the books made available for download, but also increases their file size and rendering requirements (Leetaru 2008). The question of what priority LSDIs place – or ought to place – on producing archival-quality scans will be an important one going forward. And third, the OCA has chosen to pursue a completely different model for dealing with the issue of copyright: where Google’s strategy has been to scan everything regardless of copyright status, and then to offer different levels of access to the scans based on that status (placing the burden on content owners to opt out if they object) (Google 2011a), the OCA/IA strategy (in line with decades of history in library digitization) was at least initially to confirm public domain status or secure permission before scanning materials (Leetaru 2008). The OCA’s goal, like Google’s, was ultimately to scan every book ever published, but this copyright policy, combined with other differences in administration and financing, enforced a much different order and pace of digitization. Though there is some question as to whether the OCA continues to exist *per se* (interviewees were of widely varying opinions), many of the scanning partnerships it helped to establish between the Internet Archive and book-collecting institutions continue, and new scans continue to be made available via the IA Text Archive, Open Library (an IA project), and other sites.

1.2 Early American Public Libraries

On the public library side, there are many more potential cases to choose from. The public library movement promoted the dispersion of thousands of libraries across the U.S. over the course of the nineteenth and early twentieth centuries, and at least dozens of those could be said to have influenced the movement as a whole in various ways (Ditzion 1947, Shera 1949). Nevertheless, two cases within this enormous field stand out as especially informative regarding the fundamental foundations of the movement: (1) the Boston Public Library and (2) the Carnegie Library system. These cases are notable for quite different reasons, explained in turn below.

1.2.1 The Boston Public Library

The Boston Public Library (BPL) was not the first American public library, any more than Google was the first entity to start scanning books, but it is nonetheless among the most historically significant public libraries in the U.S. – possibly the most significant.¹⁹ The

¹⁹ Another contender for this title might be Benjamin Franklin’s Library Company of Philadelphia: founded in 1731, it certainly anticipated certain features of the public library movement long before that movement truly gained a foothold, and Philadelphia at the time also held a position of influence in the emerging nation (Shera 1949, 32).

importance of the Boston Public Library derives first from its scale and location, and second from the model for emulation it provided. As of the 1840s, Boston was the fourth-largest city in the United States, and held a position of great social and economic influence nationally (Shera 1949, 170-71). Thus, when Boston's city government (and Massachusetts's state government) signaled its intent to found a public library to serve all the city's populace at that time, it served as a strong indicator of the foothold that the concept of the public library had by then attained (Shera 1949, 170). The Boston Public Library would be the first public library to serve a major urban area – and the added complexities of the urban setting (scale, economic diversity, etc.) meant that most of its structures for administration, funding, and access had to be created from scratch; it could not simply extend those of earlier, smaller public libraries. And given Boston's status as a cultural role model at the time, the library's founders anticipated – correctly as it turned out – that whatever policies and structures they put into practice would be widely emulated (Shera 1949, 179-81). Boston thus became a leader and exemplar for the entire public library movement from that point on: as Ditzion suggests, “heavily supported by public spirit and municipal aid, [Boston] seized the leadership [of the movement] and became a model both for general public policy and technical library practice” (1947, 13). In this leadership role, the original trustees of the Boston Public Library – and especially George Ticknor and Edward Everett – did much of the heavy lifting with regard to conceptualizing what public libraries could and should be from that point forward. Indeed, Shera goes so far as to suggest that the trustees' 1852 report:

still stands as the best single statement of the relation of the library to the social order. What was said then has been repeated many times since, but seldom with equal clarity and precision. In the words of Ticknor and Everett, Boston was expressing clearly the totality of public library motivations prior to 1850 (1949, 181).

Moreover, the Boston Public Library, once established, soon became one of the elite libraries in the United States, and remains so even to the present. By the late 1860s, the BPL was already the second-largest library in the United States by collection size, and, as a newspaper noted at the time, was “certainly inferior to none in usefulness or in the suitable selection of its books” (“The Experiment of Free Libraries” 1867). As a system, the Boston Public Library continues to hold the second-largest collection in the country, trailing only

However, this case has been excluded for two main reasons. First, as noted earlier, the Library Company was not a true public library, but rather a subscription library, formed primarily to serve the needs of a small discussion society known as the Junto (Shera 1949, 32-33). And second, the institution's lasting influence on the public libraries as a whole is somewhat questionable; indeed, the Library Company as such was dissolved in 1856, just as the public library movement was beginning to pick up steam (Shera 1949, 33). In fact, Shera even questions its influence on the *social* library movement to which it more accurately belonged, asserting that, “to say that this library was in any way responsible for those [other social libraries] that followed it so closely is to stretch credibility beyond reasonable limits” (Shera 1949, 56).

the Library of Congress: in fact, its 23.5 million volumes make it roughly 1.5 times the size of the next-largest library collection, that of nearby Harvard University (American Library Association 2008). The relative scale of the Boston Public Library, combined with its powerful influence on the shape of the public library phenomenon, make it an ideal case for inclusion in this study.²⁰

1.2.2 The Carnegie Library Program

The rationale for including the Carnegie Library program in this study differs somewhat from the rationale for including the BPL. Like the Boston Public Library, the Carnegie Library program was extraordinarily influential in the realm of library structure and policy. However, it accomplished this influence more through geographic ubiquity than through emulation (though the Carnegie Libraries were certainly emulated also, especially in terms of architecture). Andrew Carnegie's program of library donations spread 2,509 public libraries across the English-speaking world in the late nineteenth and early twentieth centuries – 1,679 of them in the United States (Lorenzen 1999). And especially as the program developed and matured, these library donations became much more systematized and uniform: the policies for local tax support, the priorities for building design, and the general philosophy of public access were, if not centrally planned by the Carnegie Corporation, then at least strongly shaped by the Corporation's recommendations (Van Slyck 1991, 369, 76-81). In fact, under the direction of James Bertram after the turn of the century, the Carnegie Corporation began including a set of design recommendations – including six standardized floor plans – with every library donation offer (Van Slyck 1991, 376-77). This centralized structural influence helped to institutionalize decades' worth of agitation for library policy reform: the later Carnegie Libraries show a clear trend toward eschewing extravagant architectural ornamentation; deemphasizing hierarchical relationships between donors, trustees, librarians, and users; and making the library maximally useful and usable for patrons (for example, by implementing open stacks and reducing the proportion of building space reserved for library staff and trustees) (Van Slyck 1991, 376-77). Though the Carnegie Corporation stopped making library donations in 1917, by spreading libraries throughout the country whose structure reflected particular values and ideas about library administration, the program assumed sufficient influence to make it an important model to consider in this study.

That said, however, it would be impractical to attempt a study of the entire Carnegie Library program in this context: 1,679 exemplars is simply too many. Thus, this case example will be bounded a bit differently from the other three. Specifically, since the focus

²⁰ A further factor of interest, though not a decisive one, is the Boston Public Library's current stance on large-scale digitization: in 2007, it was one of the libraries to most vocally reject Google's offer of a digitization partnership, opting instead to join the Open Content Alliance, and thus "remain open" (Albanese 2007b).

of this dissertation is on the initial negotiation of structures, the case study of the Carnegie Libraries will be comprised of two pieces: first, an examination of those features of the planning and ideation that were centralized (such as Carnegie's writings on philanthropy and Carnegie and Bertram's role in administering the donations), and second, a more detailed examination of the negotiation processes behind the earliest libraries funded by Carnegie (before or during 1899): Allegheny City, PA (1886), Pittsburgh, PA (1890), Johnstown, PA (1890), Fairfield, IA (1892), Braddock, PA (1895), Atlanta, GA (1898), Carnegie, PA (1898), and Pittsburg, TX (1898), as well as the 28 donations made in 1899. While the libraries founded prior to 1898 were not subject to the systematization that characterized the program in its later years, they certainly helped to inspire it, through their frequent cost overruns, repeated requests for additional funds, and overemphasis on the personal relationship to the donor (Van Slyck 1995). Exploring these early cases, as well as Carnegie's transition from this "retail" mode of giving to the "wholesale" mode he would pursue starting in 1899, provides a useful window into how and why the program developed in the way that it did as it progressed in later years. Additionally, because there has been a great deal more written about the Carnegie Libraries than about any of the other three cases, and because primary source data for this case is spread out among many institutions which were not all pragmatically accessible during the research period, secondary sources will play a greater role in fleshing out the discussion of the research questions for this case than for the others.

2. Data Collection

The overarching methodological structure for this comparative case study analysis is shown in the figure at right. In particular, it shows that two methodological streams were pursued: one for the two public library cases, and one for the LSDI cases. And because certain elements of the strategy for studying the LSDIs emerge from that for examining the public libraries, this section will proceed in the reverse order of the one above: it will first describe the data collection procedures for the Boston Public Library and Carnegie Libraries, then those for Google Books and the Open Content Alliance.

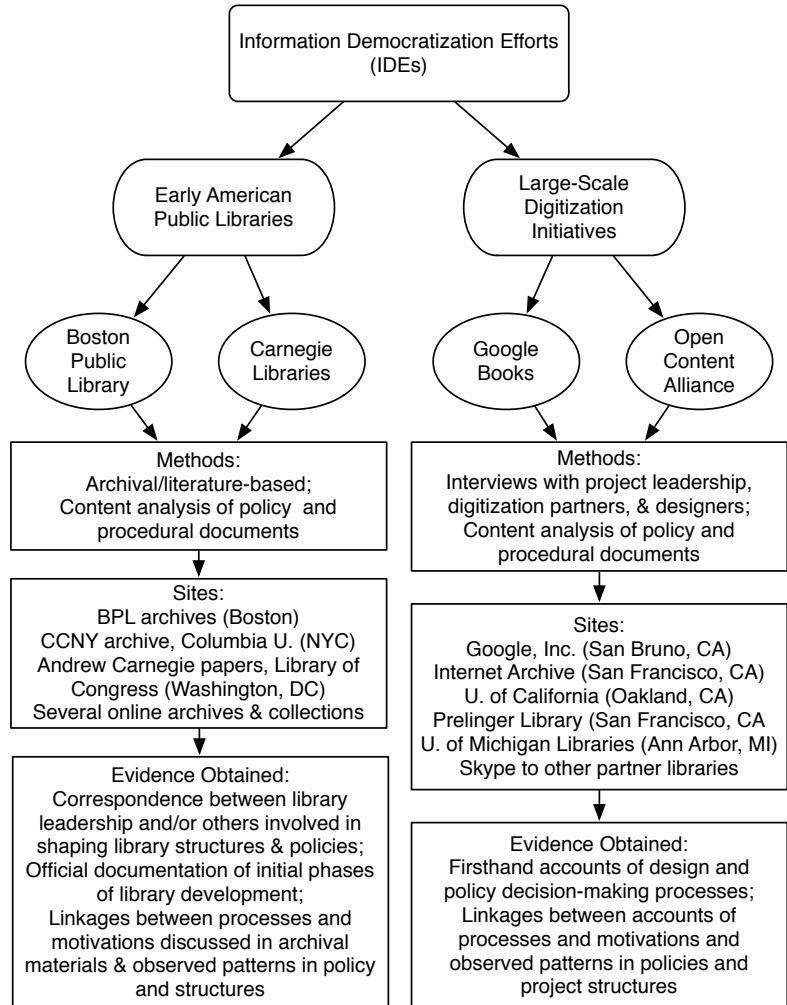


Figure 8: Data Collection Flow

2.1 Early American Public Libraries

For the early library cases, since the aim was to tease out the initial processes through which their physical and political structures developed, and those processes occurred from roughly 1840 to 1900 (Ditzion 1947, Shera 1949, Van Slyck 1995), the evidence sought has been overwhelmingly historical and archival in nature. Much of this research has been done remotely, through use of secondary sources, republications of primary source material (e.g., "Report of the Trustees of the Public Library of the City of Boston" 1852, Carnegie 1889a, Van Slyck 1995), and primary source material that has been digitized and placed in library databases or on the open web.²¹ In order to obtain the additional primary source material needed to truly illuminate the processes involved, however, time was also spent in the institutional archives for both of these library programs. The relevant materials for the BPL,

²¹ Interestingly, this remote primary source research is possible to a large extent due to the work of the very large-scale digitization initiatives being studied here: a vast number of nineteenth-century magazines, books, reports, and other resources related to both BPL and the Carnegie Libraries have already been made freely available through Google Books, the Open Content Alliance, or both.

including architectural plans, policy documents, historical newspaper clippings, and trustees' papers (correspondence, internal memos and ephemera, etc.), were retrieved mainly from the Rare Books & Manuscripts, Fine Arts, and Government Documents collections in the library's central branch at Copley Square. Some analogous materials for the Carnegie Library program were retrieved from the Rare Book & Manuscript Library at Columbia University, as well as the Manuscript Collections at the Library of Congress.²² Detailed plans for which materials to request and peruse were prepared in advance of visits to each of these archives, based on findings from online and secondary sources, as well as consultation of online finding aids and correspondence with the collections' respective librarians and/or archivists.²³

2.2 Large-Scale Digitization Initiatives

The evidence collected for the two contemporary cases is somewhat more diverse, as their development processes are much more recent – even, to a certain extent, ongoing. For these cases, the evidence provided by secondary sources such as academic literature and journalistic accounts can be supplemented by the much richer data to be gained from direct interviews with the project's leaders and designers and, to some extent, casual observations of the institutional contexts in which the projects are managed.

Although collecting additional forms of data for these two cases reduces the direct comparability of the evidence across the four cases, I suggest that doing so here is crucial due mainly to two deficiencies in the available documentation of the contemporary cases. First, correspondence among LSDI leadership occurred almost exclusively via email, and both my prior experience with these cases and my interviews with project leaders suggest that much, if not most of this correspondence has already been deleted – it is simply no longer available to draw upon. And second, even where other forms of documentation may still exist within the institutions involved, most of that documentation remains unavailable to outside researchers. This is especially true in the GBLP case, where non-disclosure agreements (and uncertainty over non-disclosure agreements) limit what can legally be shared by project participants. And further, beyond these gaps, I would also suggest that there is a more positive reason for engaging in interviews for these cases: that is, the targeted firsthand accounts to be gathered this way are simply so much richer than what can be obtained from documentary and secondary accounts that were not created

²² There are also other caches of documents on the Carnegie Library program which I was unable to get to, including those at Carnegie Mellon University and at some of the individual Carnegie Libraries. Partially as a result of this dispersion, the primary sources collected for this case are more patchy, and the analysis of the case thus relies more heavily on secondary sources.

²³ The document gathering at the Boston Public Library and at Columbia were executed exclusively by myself, in June 2011. Data gathering at the Library of Congress was executed by three individuals: two UW MLIS students each volunteered to make targeted visits to the collection while in the area on other business, and later I made my own more inclusive survey of the collection.

specifically to serve the needs of this research. Had it been possible to interview the leaders of the BPL and Carnegie cases, I would happily have done so; and yet, the absence of that possibility in those cases does not seem a sufficient rationale to ignore this rich source of data for the ones where it remains available.

The interviews conducted for this study were semi-structured in nature, with three sets of questions that proceeded in sequence. The first set of questions consisted of pure demographics: easy to answer questions such as age and educational background, meant to ease the participant into the interview, and also provide a sense of where they were coming from. The second set of questions asked the participant to guide the interviewer through the history of the project in which they were involved, from their perspective: how they became involved, what their role was, what the major milestones were. This autobiographical/event history piece served to jog participants' memories of the projects' early days, which were between four and eleven years old at the time of the interviews, depending on the participant. Such strategies have been shown to improve recall of past events in survey contexts (Belli 1998), and I believe that it was similarly helpful as a memory aid here. The third and final portion of the interview protocol delved into the more holistic questions posed: why their institution got involved, what sort of priorities existed in scanning the collections (if any), what kind of individuals they envisioned using the end product, and which individuals were especially influential in shaping the project. Before each interview, the questionnaire was customized to the participant based on their institutional affiliations and the digitization project(s) with which they had been involved – Google Books, the OCA, or both. The questionnaire was also supplemented during the interviews by open-ended follow-up questions, guided by the topics that organically arose in conversation. A generic form of the questionnaire appears in Appendix C.

The first round of participants was selected based on the researcher's prior knowledge of the project, as well as recommendations from two key individuals – one involved with Google Books, one with both Google Books and OCA. Subsequent participants were recruited using a snowball sampling strategy, based on the responses to the following two interview questions:

- *Who do you think has played an especially strong role in shaping the project? How so?*
-and-
- *Is there anyone you think it would be especially important for me to interview for this study?*

In total, 91 individuals were mentioned in response to these questions (52 for Google Books, 39 for OCA). Of those, 32 were mentioned by more than one participant (17 GBS, 15 OCA). Lawyers (two, both GBS) were excluded, in keeping with the orientation of this study toward non-legal issues. Also excluded were individuals who: (a) seemed too tangential to

have had an impact on decision-making processes, (b) would be near-impossible to reach (e.g. Larry Page), (c) were important later in the project, rather than at the beginning, or (d) represented institutions where one or more interviews had already been conducted (unless there was a compelling reason to do multiple interviews at that institution, as, for example, at Google and the Internet Archive).

All told, of the 32 individuals mentioned by more than one person, 19 were contacted, and 12 were interviewed. Additionally, ten individuals mentioned by one person or fewer were contacted, either because they were selected in the first round, or because they were incorporated later to increase the diversity of institutions and job functions represented; six of these individuals were interviewed.

Ultimately, of 29 total individuals contacted for interviews, 18 accepted and were interviewed between September 2011 and February 2012. Of these, ten were conducted in person (seven in the San Francisco Bay Area, three in Ann Arbor, MI), and eight via Skype. Interviews lasted between 30 and 90 minutes, with most taking roughly one hour. All interviews were audio-recorded and transcribed, and all transcripts were sent to the relevant participants for review. No significant alterations in transcript content were requested by any participant, though several requested minor corrections (e.g. proper spellings of names, clarifications of comments). The institutions represented by the interviews include:

Google Books	Open Content Alliance
Google (2)	Internet Archive (5*)
University of Michigan (3*)	University of Toronto (1)
Harvard (1)	Woods Hole Oceanographic Institute (1)
Stanford (1)	California Digital Library (2*)**
Oxford (1)	New York Public Library (1)**
New York Public Library (1)**	Prelinger Archive (1*)
California Digital Library (2*)**	
Committee on Institutional Cooperation (1*)	

* Individuals counted here had multiple institutional affiliations, and were counted twice in this table.

** Institution participated in both projects, so its representatives discussed both projects in their interviews.

In addition to the interviews, some casual observations of work contexts and/or scanning centers were possible, at the Boston Public Library (an OCA partner), the University of Michigan Libraries (a Google partner), and at the Internet Archive. These observations were in no way formal, but nonetheless provide some contextual insight as to the atmosphere(s) in which these initiatives were conceived and went forward. Little documentary evidence was gathered during interviews, generally because it was either confidential or had already

been deleted (or had never been written down in the first place). However, a large amount of documentary evidence – including digitization contracts, “About” and “FAQ” pages, speeches and blog posts by key stakeholders, and workflow diagrams and descriptions – was gathered before and after the interviews from existing online resource pages maintained by various digitization partners involved in both projects (Google, IA, libraries) and by other interested parties.

All primary source documents that were formally coded are listed by case in Appendix B.

3. Data Analysis

The data collection procedures outlined above produced an extremely rich set of data for analysis, including:

- a. Written evidence of process for the public library cases: e.g. correspondence, meeting minutes, city ordinances, and other institutional records.
- b. Interview transcripts for the LSDI cases.
- c. Publicly available documentation of the LSDI cases, as noted above.
- d. Secondary source materials for both the public libraries and the LSDIs.
- e. Field notes and memos documenting observations of LSDI management contexts.

These data, in turn, can be broken down into two broad categories, corresponding to the type of analysis to be applied to each:

1. Primary Source Texts: Texts emerging directly (verbatim) from a project participant (a, b, & c above).
2. Secondary Source Texts: Texts that I or another outside observer produced to explain or describe some aspect of the cases (d & e, above).

For the primary source texts, the coding and analysis proceeded based on a content analytic methodology, where “content analysis” is understood fairly broadly, as explained by Michael Patton:

Content analysis usually refers to analyzing text (interview transcripts, diaries, or documents) rather than observation-based field notes. **More generally, however, content analysis is used to refer to any qualitative data reduction and sense-making effort that takes a volume of qualitative material and attempts to identify core consistencies and meanings.** Case studies, for example, can be content analyzed (2002, 453, emphasis in original).

Whereas some other authors depict content analysis as a purely quantitative reduction of texts to numbers of words, phrases, and so forth (e.g., Berelson 1971, Nuendorf 2002), Patton (like some others, including Shapiro and Markoff 1997, Duriau, et al. 2007, Carlyle 2010) is more inclusive, describing a qualitative (or mixed qualitative-quantitative)

approach to the systematic analysis of not just the text's word and phrase counts, but also its latent meanings and revealed interpretations.

The clearest way to articulate precisely how these analytic methodologies were employed in the current study will be to describe the process in roughly chronological order.

3.1 Phase 1: Prior to Data Collection

Before going into the field – before visiting the archives and/or conducting interviews – a “start list” of codes was developed based on the theoretical framework, definitions, and research questions described above, to help guide investigation and analysis during data collection (Miles and Huberman 1984, 57-60). This list changed a great deal as data was collected, having been constructed with this sort of evolution in mind. Indeed, as this start list shifted in form, so did the research questions, through an iterative process that ended with the current set of three questions stated in the Introduction and coding framework attached as Appendix A.

This pre-data analysis phase served to set the stage for an analytic-inductive approach to data analysis. Unlike grounded theory (Strauss and Corbin 1997), analytic induction “begins with an analyst’s deduced propositions or theory-driven hypotheses” – not a blank slate to be populated inductively through exposure to indigenous concepts (Patton 2002, 454). More specifically, this process started from what Patton (2002, 456) – and Giddens (1984, 326) – refer to as “sensitizing” concepts or devices. As Patton explains, “Using sensitizing concepts involves examining how the concept is manifest and given meaning in a particular setting or among a particular group of people” (2002, 456). And Giddens cautions against extremes of theoretical implementation, noting that “to suppose that being theoretically informed...means always operating with a welter of abstract concepts is as mischievous a doctrine as one which suggests that we can get along very well without ever using such concepts at all” (1984, 327). For Giddens, theory should neither be a dictatorial ruler of the researcher’s thoughts, nor an absent element in setting about thinking about a research problem: rather, it should provide guidance as to emphases and focal points, arguably at all phases of research, from problem definition to interpretation (1984, 326-30, Pozzebon and Pinsonneault 2005, 1357-59). Thus, though there is no one “right” method to choose for the application of structurationist thinking (Pozzebon and Pinsonneault 2005), an analytical inductive strategy is as much in keeping with this school of thought as any, and more than some. For this study, the relevant sensitizing concepts were indicated by the initial research questions and theoretical framework (which in turn emerged from examination of structuration and STS theories and the literature in this area): for example, how the idea of “everyone” as a user is pragmatically negotiated and employed in each setting, or how Giddens’s concepts of rules and resources manifest themselves within or

across cases. This analysis process remained iterative and inductive, but did not presume to be completely free of theoretical framing at the outset.

3.2 Phase 2: During and After Data Collection

The process for analysis, as noted, was iterative and ongoing alongside data collection to the extent possible. It drew most heavily upon the practical advice offered by Miles and Huberman (1984), though the more specific guidance for content analysis put forward by Nuendorf (2002), and that for case study analysis presented by Yin (2003) also influenced the overall approach.

A systematic content-analytic coding process was applied to primary source texts. All actual utterances produced by project personnel, whether in the shape of interview transcripts, correspondence, public statements, or some other form, were assessed in this way, focusing on the actual words used, concepts and metaphors employed, etc. This occurred mainly using an Adobe Flash-based online data analysis utility called Dedoose (<http://www.dedoose.com>), using its streamlined excerpting, tagging, and retrieval features. Dedoose, at the time, required all documents to be in plain text. As such, the interviews were transcribed,²⁴ the print documents were run through optical character recognition software (ABBYY FineReader 11), and some of the shorter handwritten documents were either hand-transcribed or dictated into text using Dragon NaturallySpeaking 11. However, longer handwritten documents, such as the several hundred pages of BPL Trustees' meeting minutes and the dozens of reels of Carnegie correspondence microfilm were not amenable to either OCR or transcription. These were thus left as-is and coded by hand using an iPad app, PDF Expert. In these cases, the relevant (i.e. coded) passages were later transcribed into plain text and uploaded to Dedoose, and their codes reapplied in that new context.

While interviews were still ongoing, two sample sets of documents (two documents from each case per set) were hand-coded using earlier iterations of the code book, to test the efficacy of the codes. This coding was conducted by myself and a University of Washington Master's student, Sarah Caldwell, and through iterations of this process and conversations between the two of us, the code book was revised to its current form. This collaboration was extremely valuable because it offered a second set of eyes on the data, which, as Miles and Huberman note, can improve reliability:

Definitions get sharper when two researchers code the same data set and discuss their initial difficulties. Disagreements show that the definition has to be expanded or otherwise amended. ... [Furthermore,] Double-coding not only aids definitional clarity, but is a good reliability check. Do two coders, working

²⁴ Some by myself, some by UW MLIS student Sarah Caldwell, and most by a UW iSchool GA and MSIM student, Adam Taplin.

separately, agree on how big a codable block of data is? And do they use the same codes for the same blocks of data? (1984, 60-63).

Once the code book had reached a satisfactory point, the primary source documents were coded in roughly chronological order, from the earliest project (Boston Public) to the latest (OCA). Some jumping back and forth did occur, as new documents were located, and also simply to ease the tedium of reading meeting minute after meeting minute (or contract after contract) for days at a time. All told, 196 documents were coded: 181 in Dedoose and 15 on the iPad. These varied in length from one or two lines to hundreds of pages, with the longest being the 455-page, handwritten *Proceedings of the Trustees of the Boston Public Library*, 1852-1858. Broken down by case:

- 38 documents related to the Boston Public Library,
- 36 to the Carnegie Libraries,
- 72 to Google Books,
- 47 to the Open Content Alliance, and
- 3 (all interviews) to both Google Books and the Open Content Alliance.

As noted above, these are listed in Appendix B.

Alongside the formal coding of transcripts and other documents retrieved, I also engaged in extensive memoing about my observations and reflections at least daily during direct archival investigations and interviewing, and also frequently during analysis, reflecting upon the themes present in each document coded, as well as my general perceptions and overarching emergent themes. These memos included informal descriptions, thoughts, and reflective remarks (Miles and Huberman 1984, 64-65) and, for interviews, also a more formalized contact summary form (Appendix D) to help me to remember the more tacit or observational elements of each interaction (Miles and Huberman 1984, 54).

Secondary sources (including my own memos) were not analyzed as formally as primary source texts, as they do not reflect the direct statements of informants, but rather the already-processed analysis of another researcher or observer. Rather, they were viewed more holistically as a source of triangulation for the evidence offered by the primary source texts. This triangulation proved especially helpful for factual tasks such as pinning down the exact timelines of the two digitization initiatives, and also to check my own interpretations of the historical evidence, much of which has been analyzed previously by others, for different purposes. Much of this occurred during secondary analysis, described below.²⁵

²⁵ Also worth noting: a vast amount has been written about both the two library cases and the two LSDIs (though especially the Carnegie Libraries and Google Books). The secondary source texts relevant for inclusion here do not

3.3 Phase 3: Secondary Analysis

After all data had been collected and initially coded in Dedoose, the resulting excerpts were output to Microsoft Excel in separate workbooks for each case, and within those, separate spreadsheets for each code, and these excerpt documents were then re-coded in an emergent, grounded fashion, tracking the sub-themes that appeared within the data itself. As each sub-set of codes emerged, memos were written that extended the analysis of each primary code into prose drawing out themes and potential connections both between codes within each case and between the cases. These served as the bases for writing each of the case chapters, and ultimately, collectively, for drawing the broader conclusions presented at the end of this analysis.

represent a census of these writings, but rather a purposive sampling of those writings based on the need to confirm or contest claims made in primary sources.

Case 1: Boston Public Library (1848-1865)

In comparison to the other three cases examined here, the Boston Public Library's early history is extraordinarily well-documented. This is so for at least two reasons. First, the requirement that three annual reports be produced – one by the Trustees, one by the Librarian, and one by a partially-external examining committee – was written into the original local law that founded the institution (Boston Common Council 1852a). As official government documents, these reports remain available in a number of locations, and taken together, they provide excellent insight into the foundation, functioning, and growth of the institution. The second reason for the extensive availability of extant documentation of the library's history is its status as a basically stable, unitary institution – and one dedicated to collecting and preserving various sorts of records – over more than 150 years. Because of this history, not only have many potentially ephemeral documents – such as personal letters, meeting minutes, and potential building designs – been preserved; they are also largely still available within the bounds of a few departments at the library's central Copley Square Building: namely, Rare Books and Manuscripts, Government Documents, Fine Arts, and the general delivery desk. In addition to this treasure trove of primary source documents, moreover, two very useful histories of the library have been written, by Horace Greeley Wadlin (1911) and Walter Muir Whitehill (1956), which help to position the primary texts within a broader arc.

This examination of the early days of the Boston Public Library (BPL) will not attempt to replicate the work of those earlier historians by providing a comprehensive account of the institution's chronology. Instead, establishing the structure to be followed by the other three case chapters, it provides a brief synopsis of the institution's early history (for BPL, up to roughly 1865 – the period in which the library was established, its first dedicated building was built, and the institution within that building became both extremely popular and extremely influential), and then examines how the three research questions posed in the introduction apply to this case: in brief, (1) the expressed motivations behind the institution, (2) how the institution was defined by those influencing its design, in terms of users and collections (respectively), and (3) the processes and structures put into place in order to

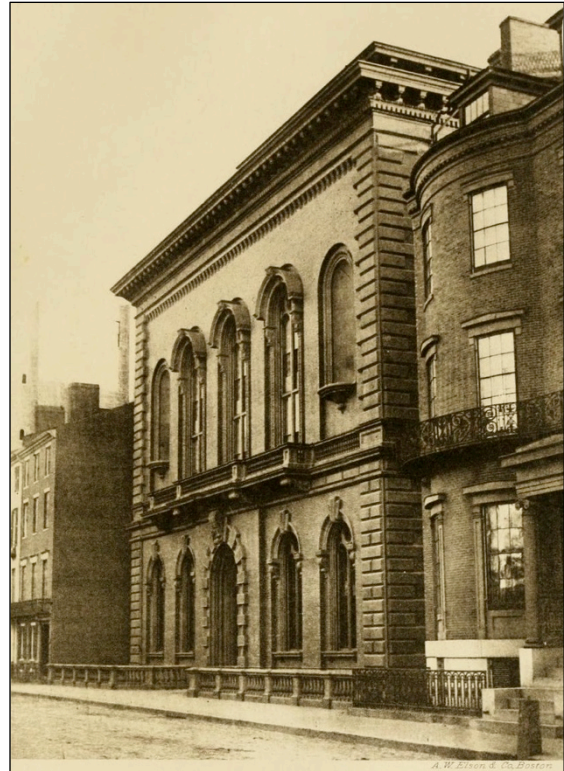


Figure 9: Boylston Street Building
Location of Boston Public Library, 1854-1895
(Image: Wadlin 1911, 66)

strive toward these motivations and definitions. Figure 10, below, provides a visual representation of the basic chronology for the period 1840-1875.

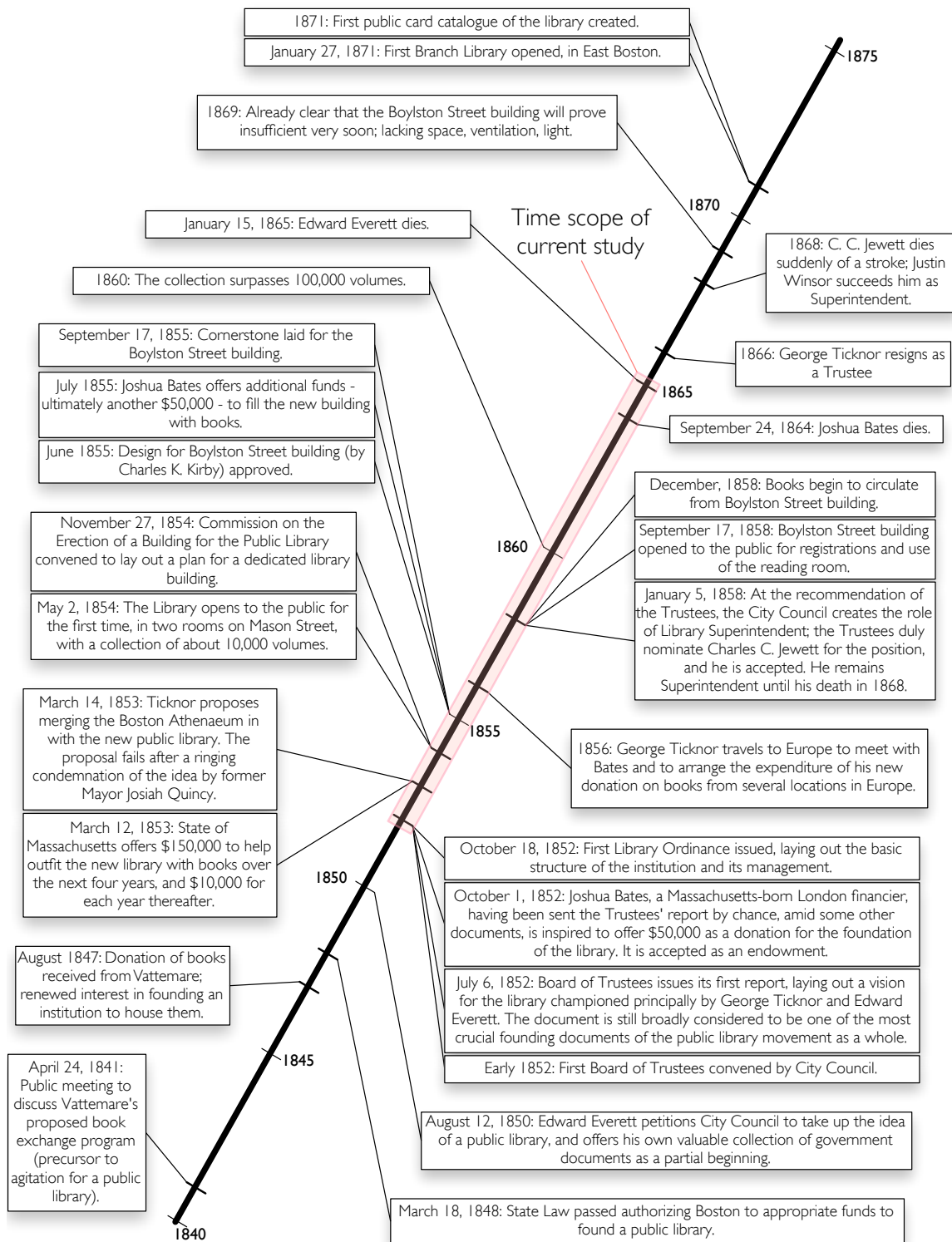


Figure 10: Timeline of Boston Public Library Development, 1840-1875

1. Historical Synopsis

Discussion of establishing a public library in Boston extends back to at least 1841, when French ventriloquist Alexandre Vattermare organized a public meeting to discuss a system of international exchanges of books, and as a corollary to that system, the merging together of fifteen existing Boston cultural institutions into a single, public-access entity (Whitehill 1956, 7). That plan failed fairly decisively, when only one of said cultural institutions expressed willingness to participate; four others politely but firmly declined, and the remaining ten didn't even dignify the idea with a response (*Documents Relating to a Meeting of Citizens of Boston Held May 5, 1841, to Consider Mr. A. Vattermare's Plan of a Public Literary and Scientific Institution [in Boston] and System of International Literary Exchange 1841*, Whitehill 1956, 8-9). Still, most accounts attest that this event planted the idea of a public library in the minds of Bostonians (e.g., Wadlin 1911, 5, Whitehill 1956, 9-11), though it would only come to fruition a decade later, and in a much different form.

By the late 1840s, agitation for a public library had increased, gaining the support of Boston Mayor Josiah Quincy, Jr. In 1848, at Mayor Quincy's urging (via the Boston City Council), the Massachusetts State Legislature passed legislation authorizing the city of Boston to establish a public library, supported at least in part by public funds (State of Massachusetts 1848). As Wadlin notes, this "was the first statute ever passed authorizing the establishment and maintenance of a public library as a municipal institution supported by taxation" (1911, 11) – quite an historic moment for the public library movement. Still, the idea languished in the City Council for a few more years, until 1852, before concrete action was taken. In the intervening years, books and donation offers began rolling in for the still-unprepared library, including two shipments of books from Vattermare, \$1000 from John P. Bigelow²⁶ (Quincy's successor as Mayor), and a large and valuable collection of public documents and State papers from Edward Everett (a former Congressman, Massachusetts Governor, U.S. Minister to the United Kingdom, and President of Harvard University – and soon to be a founding Trustee of the Boston Public Library) (Whitehill 1956, 11-23). Ultimately, it took the urging of yet a third Boston Mayor, Benjamin Seaver, to spur the City Council into taking concrete steps toward establishing the library: first, procuring a temporary location for the institution, in a few rooms of an existing city building, and second, appointing the library's first librarian and board of trustees, tasked with laying out a vision for the institution and shepherding it into reality. Both of these occurred in May of 1852 (Whitehill 1956, 25-26).

By most accounts, the activities of the Board of Trustees – and of the library as a whole – for the first several years were shaped by just two individuals: George Ticknor and Edward Everett. Already personal friends by 1852, the two men had been discussing their ideas for

²⁶ The equivalent of about \$27,000 in 2013 dollars.

a public library in Boston for years by the time they were named as founding library trustees in 1852 (Whitehill 1956, 23-24). In fact, Ticknor had been dreaming of such an institution for decades by then, and had very specific ideas about how it should be run. In particular, Ticknor made clear that he would have no part in founding the institution if the books were not allowed to freely circulate – a point upon which he differed with Everett, but which Everett graciously conceded (Whitehill 1956, 26). Recalling his stance in a letter to the Trustees nearly a decade later, Ticknor wrote:

From the earliest suggestion of such an institution, it has been my prevalent desire that it should be made useful to the greatest possible number of our fellow-citizens, especially to such of them as may be less able than they would gladly be to procure pleasant and profitable reading for themselves and for their families. This is known to all the Trustees with whom I have successively served, and our President remembers, that I never would have put my hand to the Institution at all, except with this understanding as to its main object and management. Nor has there been any real difference on this point among the different persons who have controlled its affairs during the eight years of its existence ("Eighth Annual Report of the Trustees of the City Library" 1860).

In July, the Trustees issued their first report to the City Council, virtually entirely written by Ticknor and Everett ("Report of the Trustees of the Public Library of the City of Boston" 1852).²⁷ As noted in the introduction, it was a masterwork, frequently lauded in the years since as one of the foundational documents of the public library movement as a whole (e.g., Shera 1949, 181). Other expressions of the motivations of those responsible for establishing the Boston Public Library do exist and will be drawn upon during the discussion of that topic below; however, were this report the only resource for answering that question, it could give a nearly comprehensive view. The report called for the creation of a public, tax-supported, circulating library, geared toward filling a perceived gap in the then-new public education system by providing the “means for general popular instruction and self-culture” – that is, freely available books and other information resources – to those who had completed their secondary education and did not have the financial means to acquire or gain access to such resources in other ways (that is, by purchasing either the materials themselves or a membership in an institution that might provide them, such as the Boston Athenaeum) ("Report of the Trustees of the Public Library of the City of Boston" 1852, direct quote from 14). It also laid out four types of materials that such a library might be expected

²⁷ In fact, it is possible to tell without much difficulty which of the two men wrote which piece, as many sections are cribbed directly from their letters to one another. Ticknor's biographer explains that, when the two men were appointed as a sub-committee tasked with drawing up the report, "Mr. Ticknor prepared...a paper, expounding on the principles and plan on which the institution was to be founded, -- these being his own, -- and Mr. Everett left this *entirely untouched*, adding some pages, at the beginning and end, on the general import of the project" (Hillard, et al. 1876, 305). Specifically, Ticknor wrote pages 9-21, and Everett the rest. A copy of a version of the report annotated by Charles C. Jewett, the BPL's first Superintendent, specifying these distinctions, was included as an appendix to Jesse Shera's classic *Foundations of the Public Library* (1949, 267-90).

to make available – non-circulating reference works, circulating books requiring only a single copy, circulating books requiring many copies, and periodicals – and requested funds for their acquisition, as well as the hiring of the Librarian and the arrangement of rooms in a city building to give the library a start ("Report of the Trustees of the Public Library of the City of Boston" 1852, 16-19, 21-23).

Interestingly, the 1852 Trustees' Report turned out to have not only great descriptive power, but great power of persuasion. In that same year, the city of Boston was in the process of negotiating a water loan with a London financial firm called Baring Brothers & Company, and their contact person on this loan was a Massachusetts-born financier named Joshua Bates. In negotiating the loan, a sheaf of City Council documents was sent for Bates to review; among these documents, but in no way relevant to the matter at hand, was the Trustees' report (Greenough 1891?, 5-6). This tangential inclusion would prove fateful for the future of the library. On October 1, Bates wrote to Mayor Benjamin Seaver:

I am indebted to you for a copy of the Report of the Trustees of the Public Library for the City of Boston, which I have perused with great interest, being impressed with the importance to the rising and future generations, of such a library as is recommended, and while I'm sure that, in a liberal and wealthy community like that of Boston, there will be no want of funds to carry out the recommendation of the Trustees, it may accelerate its accomplishment, and establish the library at once on a scale that will do credit to the city, if I am allowed to pay for the books required, which I am quite willing to do – thus leaving to the city to provide the building and take care of the expenses.

The only condition I ask is, that the building shall be such as shall be an ornament to the city – that there shall be room for one hundred to one hundred and fifty persons to sit at reading tables – that it shall be perfectly free to all, with no other restrictions than may be necessary for the preservation of the books. What the building may cost I am unable to estimate, but the books – counting additions during my lifetime – I estimate at \$50,000, which I shall gladly contribute, and consider it but a small return for the many acts of confidence and kindness I have received from my many friends in your city (Trustees of the Public Library 1865, n.p.).²⁸

This massive donation fueled the expansion of the ideas in the Trustees' report into concrete reality, both directly, as it provided the first stable endowment fund for the purchase of library materials, and indirectly, as it provided an enormous demonstration of faith in the project at a very early point, which helped to spur action on the part of both the city government and other donors to the library (Wadlin 1911, 44-46).

In 1854, the library opened to the public for the first time, operating out of two small rooms in the old Adams School building on Mason Street, with the interior room used to store the

²⁸ This offer would be equivalent to about \$1.4 million in 2013 dollars.

collections and provide limited space for staff, and the outer room designated as the Reading Room (Whitehill 1956, 43). Two things were clear almost immediately: first, the library would be immensely popular – in the first five and a half months alone, 6,590 people applied for library cards, and borrowed a total of 35,389 books ("Second Annual Report of the Trustees of the City Library" 1854, 16-17) – and second, the space was far too small. In fact, the latter was clear even before the rooms opened to the public; the Examining Committee noted in its 1853 report that "the accommodations in Mason street will be too small; and before a proper building can be erected, they will be found insufficient even for putting on convenient shelves the books that will have been received" ("Report of the Committee to Examine the Library" 1853). This expectation was in every way confirmed. The next year, the Examining Committee decried the conditions in the library rooms, declaring the Reading Room "noisy, uncomfortable and unfit for its peculiar purposes as a place for quiet reading; – besides which it often happens, that the persons who come to borrow books are so numerous as to be most inconveniently crowded together; many of them being kept standing and waiting much longer than ought to be necessary, until their turn comes round to be served," and the interior room "small, ill-lighted, ill-ventilated, cold in winter, and so nearly filled with books that it will soon be impossible to find places for more" ("Report of the Examining Committee" 1854, 13). They therefore went on to assert:



Figure 11: The Commission on the Erection of a Building for the Public Library, ca. 1855. Photo courtesy of Boston Public Library, on Flickr.

The case is a plain one. Every body who visits the Library, and the thousands who use it, know and feel every time they come there, that its present rooms are wholly insufficient for its present wants, – to say nothing of the increasing wants and inconveniences that must arise from the increased numbers who are constantly demanding its privileges and benefits for themselves and their families It is obvious, therefore, that something should be done with out delay for the erection of a proper building; – a building becoming the character of the City, and becoming the pledges it has given ("Report of the Examining Committee" 1854, 13-14).

Eventually these complaints bore fruit, in the form of a Commission on the Erection of a

Building for the Public Library (Figure 11²⁹), convened by the City Council in November 1854. By 1855, the Commission had selected a design by Charles K. Kirby, to be placed on a lot on Boylston Street, facing the south edge of Boston Common (Whitehill 1956, 45-47). This building was duly erected, and opened to the public on September 17, 1858.

While the Boylston Street building was under construction, however, Joshua Bates transformed the library's financial situation once again, offering *another* donation of \$50,000, this time to fill the new building with books. As Bates wrote to Everett on July 6, 1855,

It never will do to open new building with empty shelves. The fund [that is, the earlier Bates endowment] will not work fast enough. The only fear I have ever had about the library has been that you would not be able to start on the scale I should wish, viz., that would do credit to the city, and to which the inhabitants would point with pride in all future times (Trustees of the Public Library 1865, n.p.).

This second massive influx of funds, paired with the donor's expressed desire to have them expended before the new building opened, led George Ticknor himself to make an extended trip to Europe in 1856-57 to meet with Bates, purchase as many books as he could himself, and make arrangements with purchasing agents in cities across the continent.³⁰ As a result of this expedition, the library collections more than doubled in a single year, from 34,896 volumes in 1857 to 70,851 in 1858, with 24,618 coming solely from Bates's second donation.³¹

Preparing these thousands of new books for public use was a massive feat, and one taken on by another exceptional individual, Charles C. Jewett, who had joined the BPL to catalogue the Bates donation after resigning as Librarian of the Smithsonian Institution in 1855 (Whitehill 1956, 48-49). Shortly before the new building was to open in Boston, the BPL Trustees petitioned the City Council to create the office of Superintendent of the library, in recognition of the facts that (a) running an institution on the scale they desired would require a fairly impressive skill set, and (b) they already had in their employ such a skilled individual – Jewett – but he was not the City-Council-appointed Librarian, and was not likely to become so (that position had been filled by the serviceable, but by all accounts fairly unremarkable, Edward Capen, since the library's founding, and he showed no signs of leaving at that point). The Trustees also asked that the position be treated like one with academic tenure: that is, free of “the uncertainty of an annual election, by public bodies

²⁹ All photos credited to the Boston Public Library were retrieved from their Flickr page, unless otherwise noted (http://www.flickr.com/photos/boston_public_library).

³⁰ While in Paris, Ticknor also met with Vattemare – however, like so many interactions between that gentleman and the library, that meeting ultimately proved fruitless (Whitehill 1956, 61).

³¹ Though the books were purchased largely in 1856 and 1857, it took a great deal of time and labor to accession them into the library's collections, and thus into the library's statistics; thus, they appear on the 1858 report rather than the 1857.

partaking largely of a political character,” or, more plainly, out of reach of the City Council, and under the control of the Trustees themselves (Board of Aldermen 1857). Though the Council did not grant them this control over the appointment, it did allow for the creation of the role of Superintendent, as well as the effective demotion of the Librarian (Capen) to what we would now call the Head of Circulation, in order to make way for Jewett. Jewett held the position exceedingly successfully – making innovations in cataloging, circulation, and stacks management – from 1858 until his death in 1868, when he suffered a stroke while at work at the library (Whitehill 1956, 54-75).

Much further than this in the history of the institution would go beyond the scope of this study; however, for context, I will note a few later points of interest. First, the Boylston Street building, the library’s location since 1858, was already proving insufficient by 1867. The Examining Committee for that year – headed by Justin Winsor, who would succeed Jewett as Superintendent in the following year – decried the lack of light and ventilation, the difficulty of retrieving books from the upper ranges of the book alcoves, and the lack of space for library workers to perform tasks such as cataloging and preparing books for shelving (Whitehill 1956, 68). This time, however, the process for siting, designing, and erecting the building took considerably longer, and the new edifice in Copley Square, designed by McKim, Mead, & White – which the Boston Public Library still inhabits today – did not open to the public until nearly thirty years later, in 1895. In the intervening years, virtually all the individuals who were instrumental in the founding years – Everett, Bates, Ticknor, Jewett – died. Winsor proved a magnificent leader for the institution – perhaps even more accomplished than Jewett. Under his tenure, the library opened six branches, developed its first public card catalog, and began to open its doors on Sundays for the first time (among many other innovations) (Whitehill 1956, 75-101). However, a petty dispute over Winsor’s salary within the City Council, in conjunction with a more stable job offer from Harvard, led to Winsor’s reluctant resignation from the library in 1877 (Whitehill 1956, 104-08). After his departure, the institution entered what one of its chroniclers would call “the doldrums:” no comparably passionate and able Superintendent could apparently be found, and thus none was appointed. As a result, until 1895, the library ended up being run by



Figure 12: Construction of the McKim Building, Copley Square, August 1890. Photo courtesy of the Boston Public Library, on Flickr.

committee – in the form of the Trustees, who had gained greater autonomy from the City Council in the wake of the Winsor debacle – moving away from its initial mission, and losing much of its shine as an international beacon of public library aspirations for many years (Whitehill 1956, 111-30). Though things would begin to look up after the opening of the Copley Square Building and the hiring of Herbert Putnam as Librarian in 1895 (Whitehill 1956, 130), this synopsis will end there, and leave the tale of what happened next to other tellers.

2. Motivations

Many individuals and groups – from Ticknor and Everett to a series of Boston mayors and city councilmen to the institution’s early staff – exerted influence over the initial shape and direction of the Boston Public Library, and each acted based on their own particular constellations of motivations, deployed at varying levels of conscious recognition, and with varying degrees of impact. However, among the extant writings of these individuals, a relatively concise set of common spurs to action emerges. In particular, three primary motivations rise above the rest: (1) supplementing and extending the local system of public education, (2) creating a more egalitarian social playing field by providing opportunities for self-driven uplift, and (3) providing an atmosphere and resources conducive to the elevation of morals and development of better character. Further, three secondary motivations, though not as ubiquitous within the documentary evidence, also emerge as useful to understanding the eventual shape of the institution: (1) increasing Boston’s municipal prestige, especially as compared to nearby New York City; (2) supporting regional economic development by providing for the information needs of tradesmen and other skilled workers; and (3) meeting a “public demand” that was initially more of an assumption than a proven reality. Most of these motivations – both primary and secondary – share a unifying underlying thread of paternalism, though of a generally benevolent sort. This section will describe each of them in turn.

2.1 (Adult) Education

As of the mid-nineteenth century, Boston had already developed a remarkably complete system of public schools, extending from age four through adulthood, and including both boys and girls – though not to equal extents (Boston School Committee 1949, 15-18). This system was an unmistakable point of pride for the individuals tasked with laying out the functions to be served by the new public library: as the Joint Standing Committee on the Public Library boasted in 1852,

It is believed that the City of Boston expends a larger sum, annually, for the education of her children, in proportion to the number of inhabitants, than any other city in the world (Boston Common Council 1852b, 3).

Yet, this committee, as well as the original trustees, the standing mayor, and the

institution's most prominent donor, Joshua Bates, suggested that this system remained incomplete, because the learning stopped at the point where the youth left school, because many of them lacked the financial wherewithal to gain access to further educational materials (which at the time were generally housed in either private, subscription-based libraries or private collections, or offered for sale by booksellers). The same page of the report just cited went on to contend as much, and the following month, Mayor Benjamin Seaver echoed the claim, declaring that: "A Free Public Library may justly be regarded as necessary to the completion of that noble system of Public Instruction, which reflects so much honor upon our City, and does so much to promote its prosperity" (Seaver 1852, 4). These sentiments were then expanded upon at length in the founding Trustees' Report issued a few months later – indeed, they occupy several pages. Without duplicating these sections in full (though their eloquence makes it tempting), the following passage³² provides an indication of their contents:

Although the school and even the college and the university are, as all thoughtful persons are well aware, but the first stages in education, the public makes no provision for carrying on the great work. It imparts, with a noble equality of privilege, a knowledge of the elements of learning to all its children, but it affords them no aid in going beyond the elements. It awakens a taste for reading, but it furnishes to the public nothing to be read. It conducts our young men and women to that point, where they are qualified to acquire from books the various knowledge in the arts and sciences which books contain; but it does nothing to put those books within their reach. As matters now stand, and speaking with general reference to the mass of the community, the public makes no provision whatever, by which the hundreds of young persons annually educated, as far as the elements of learning are concerned, at the public expense, can carry on their education and bring it to practical results by private study ("Report of the Trustees of the Public Library of the City of Boston" 1852, 6-7).

As such passages indicate, the public library was seen as filling a gap in education *for adults*, or at least for individuals who had already completed their course of publicly-offered secondary education. The library would provide the raw materials for individuals beyond school age to continue their education in a self-directed way, in keeping with the popular nineteenth-century philosophy of "self culture"³³ (Ditzion 1947, 54, Shera 1949, 226-28, Garrison 1979, 42). Serving the needs of those under age 16 is scarcely mentioned in these early years, though doing so would later become a central role of public libraries, including Boston's. The early leadership of the Boston Public Library was strongly oriented toward education, but in their minds, it was education at a fairly high level – the assumption being that the lower levels were more than adequately provided for in the schools.

An additional theme underlying these educational motivations – and the particular, adult-

³² Likely in Everett's voice.

³³ A linkage explicitly made by Ticknor in one of his early letters to Everett (Hillard, et al. 1876, 301).

oriented form they took – was the idea that the library, by facilitating the increased intellectual attainment of voting-aged citizens, would help to improve the functioning of the democratic society in which those citizens participated. For example, the library’s founding ordinance reminds its readers that “the founders of our Republic have left on record their testimony that the perpetuity of our institutions depends upon the intelligence of the people” (Boston Common Council 1852b). And the founding Trustees’ report makes the connection still more plain, contending:

For it has been rightly judged that, – under political, social and religious institutions like ours, – it is of paramount importance that the means of general information should be so diffused that the largest possible number of persons should be induced to read and understand questions going down to the very foundations of social order, which are constantly presenting themselves, and which we, as a people, are constantly required to decide, and do decide, either ignorantly or wisely. That this can be done, – that is, that such libraries can be collected, and that they will be used to a much wider extent than libraries have ever been used before, and with much more important results, there can be no doubt; and if it can be done anywhere, it can be done here in Boston ("Report of the Trustees of the Public Library of the City of Boston" 1852, 15 (marked 17)).

The library, thus, in its educational role, was intended to increase the wisdom of the Boston voter as a participating member of the democratic system. Though the question of whether it actually served that purpose is a difficult one to answer (and no such feat will be attempted here), subsequent sections of this chapter will describe how both the educational motive in general, and its particular orientation toward adult learners and civic participation, directly influenced the types of users the institution welcomed (or not) and the types of collections the institution wished to build (again, or not) – as well as the types of systems and structures put in place to serve those aims.

2.2 Egalitarianism & Uplift

The idea that a public library in Boston would serve as a great social equalizer formed a second key motivation for establishing the institution. This motivation was interwoven with, yet distinct from, the educational aim. The two dovetail most closely on the topic of democratic citizenship, with education serving as an equalizing force among citizens who, according to their nation’s founding document, ought to have been “created equal,” yet nonetheless fell into a broad range of positions on the ladder of wealth. Education, it was thought, would help to elevate the political discourse across class divides, as well as providing the diligent and worthy with the tools to strive to improve their economic and social standing. And while education was offered on an egalitarian basis up to a point – via the public schools – resources for continuing education were still largely restricted to the upper classes, locked away in private (individually-owned) or subscription- or membership-based libraries. Thus, the public library was advocated as a force for

equalization of opportunity as well as class mobility. As the founding Trustees contended,

But it is admitted, — or else another and more general library would not now be urged, — that these valuable libraries³⁴ do not, either individually or in the aggregate, reach the great want of this city, considered as a body politic bound to train up its members in the knowledge which will best fit them for the positions in life to which they may have been born, or any others to which they may justly aspire through increased intelligence and personal worthiness. For multitudes among us have no right of access to any one of the more considerable and important of these libraries; and, except in rare instances, no library among us seeks to keep more than a single copy of any book on its shelves, so that no one of them, nor indeed, all of them taken together, can do even a tolerable amount of what ought to be done towards satisfying the demands for healthy, nourishing reading made by the great masses of our people, who cannot be expected to purchase such reading for themselves ("Report of the Trustees of the Public Library of the City of Boston" 1852, 15 (marked 17)).

This bent toward equalization of opportunity also influenced the types of users the founding leadership considered especially important for the library to serve: that is, the poor and working classes (see, e.g., "Report of the Trustees of the Public Library of the City of Boston" 1852, Trustees of the Public Library 1865). This emphasis on elevating the lower classes will be discussed further in subsequent sections – as will its impact on the library's collection development priorities and the order in which particular sections of library buildings and collections were prepared for use.

2.3 Moral/character Development

The idea that the library would be a boon to public morals and the improvement of character in members of the lower classes forms a third dominant theme within the BPL's early documents. Though not quite equal in pervasiveness to the educational or social mobility themes, the theme of moral and character development sounds particularly strongly in communications from donors to the library, and especially Joshua Bates. In a letter to his friend Thomas W. Ward, accompanying his initial donation offer (which was sent to the mayor via Ward), Bates explains his reasoning thusly:

my own experience as a poor boy convinced me of the great advantage of such a library. Having no money to spend and no place to go to, not being able to pay for a fire or light in my own room, I could not pay for books, and the best way I could pass my evenings was to sit in Hastings, Etheridge, and Bliss' bookstore

³⁴ That is, the numerous other forms of library already present in Boston at the time; specifically, the Trustees list the collections "of the Athenaeum, of the American Academy, of the Historical Society, and of the General Court, the Social Library of 1792, the Mercantile Library, the Mechanics Apprentices' Library, the Libraries of the Natural History Society, of the Bar, of the Statistical Association, of the Genealogical Society, of the Medical Society, and of other collective and corporate bodies; and coming down to the "Circulating Libraries" strictly so called; the Sunday School Libraries, and the collections of children's books found occasionally in our Primary Schools" ("Report of the Trustees of the Public Library of the City of Boston" 1852, 14)

and read with a kindly permitted me to, and I am confident that *had there* been good warm and well lighted rooms to which we could have resorted with proper books, nearly all the youth of my acquaintance would have spent their evenings there to the improvement of their minds and morals. Now it strikes me that it will not do to have the rooms in the promised library much inferior to the rooms occupied for the same objects by the upper class. Let the virtuous and industrious of the middle and mechanic class feel that there is not so much difference between them. Few but worthy young men will frequent the library at first, they may draw others from place to treading the same path, and with large, well-lighted rooms, while warm in winter, I feel sure the moral effect will keep pace with mental improvement and it will be carrying out the school system of Boston as it ought to carry out (Bates 1852).

As Bates's letter indicates, the moral rationale for the library relates both to its collections and the library-as-place. The library's collections could help elevate the morals of the populace, it was thought, by excluding "unhealthy" reading and drawing individuals into a process of taste elevation (e.g., *Proceedings of the Trustees 1852-1858*, letter from Everett to a donor (July 18, 1853) and preface to the catalogue (October 9, '58), "Eighth Annual Report of the Trustees of the City Library" 1860, Superintendent's Report). And the library as place could help improve morals and character by providing a warm, well-lit "third place" outside of home or work (Oldenburg 1991) for young adults to spend time that would be more wholesome than other entertainments available to them. Still, there is a bit of tension on this point, particularly relating to usage policies, which will be discussed more fully later on: on the one hand, the library was seen as a place where individuals could improve their characters; on the other, they had to prove they were already of reasonable character to make full use of the collections (e.g. by providing references for a borrower card); there is also a sense that to some extent, only already "worthy" individuals were really welcome – indeed, the first set of library regulations, proposed in 1853, welcomed only individuals "of respectable character, and of such orderly conduct and condition as not to interfere with the occupations and comfort of others, who may be in the room at the same time" (Board of Trustees 1853). This is a tension that Garrison (1979) discusses in great detail. Nevertheless, at the founding of the institution, the idea that it would have a positive influence on the morality and character of the lower classes clearly held appeal to the generally middle- to upper-class Trustees, donors, and politicians who helped shepherd it into reality.

2.4 Secondary Motivations

Education, egalitarianism and uplift, and moral and character development emerge quite clearly as the predominant rationales for the creation of the Boston Public Library. However, as noted earlier, three less central motivations – municipal prestige, economic development, and public demand – are also worthy of comment.

2.4.1 Municipal Prestige and Competitiveness

The idea that the library would increase Boston's prestige as a municipality comes up largely, though not exclusively, in reference to New York. In 1848, John Jacob Astor had bequeathed \$400,000 to establish a free public library in New York City. Though the library was to be broadly opened to the public – Astor specified that it should be “accessible at all reasonable hours, for general use, free of expense to persons resorting thereto” (Lydenberg 1916, 559) – it was nonetheless quite a different sort of institution than that proposed in Boston, in that it received no public support and did not allow borrowing.³⁵ Still, its establishment was a compelling spur to action in Boston: as Mayor Seaver admonished in his 1852 rallying call to the City Council:

Boston ought not long to be far behind her sister City of New York, in the establishment of a Public Library; and, while we can scarcely hope to rival her princely Astor, it cannot be doubted that we have many citizens who would be ready to bestow upon it large sums in money and in books, if they can be fully satisfied of its permanent foundation and ultimate success (Seaver 1852).

Five years later, a sense of relief comes through in the Trustees' report, as it notes how, although Boston had already fallen behind New York “in population, commerce, and material resources,” the library helped to maintain the city's primacy in the cultural realm:

Had she consented to remain without a great public library, she must have submitted also to a secondary position in the means of intellectual culture. The establishment of such an institution has prevented this reproach from coming upon her, and will enable her, in this respect, to sustain a generous and mutually beneficial emulation with her sister cities (“Fifth Annual Report of the Trustees of the City Library” 1857).

Further, the issue of prestige also emerges at a more global level – the BPL's leadership and supporters, after all, saw themselves not just as trying to build the best public library in the region or the country, but as trying to build a truly world-class institution. As Bates wrote to Everett in 1854: “I hope the City Government will commence the building of large-scale, so that in due time Boston may boast of one of the best public libraries in the world” (Trustees of the Public Library 1865). The influence of this motivation on specific structures and policies is less clear than some others; however, it seems reasonable to attribute to it two more general effects: first, it spurred the city council to take concrete action toward establishing the BPL in 1852, and second, as the institution went forward, it likely played a

³⁵ A few other interesting facts about the history of the New York Public Library: the Astor Library ultimately opened to the public the same year as the Boston Public Library, 1854. However, the New York Public Library as such did not exist until 1895, when the Astor Library was merged with the Lenox Library and Tilden Trust (Stone 1977, 158), and the institution did not circulate books until merging with the New York Free Circulating Library in 1901 (New York Public Library 2012). Also interestingly, a few years later, the city would become the recipient of the single largest library grant ever approved by Andrew Carnegie: \$5.2 million for the construction of branch libraries throughout the boroughs (Jones 1997, 13, New York Public Library 2012).

role in establishing the desired scale of the institution – that is, as large as possible, in terms of collections, buildings, and usage...and ideally larger in all those ways than New York's.

2.4.2 Economic Development

Although it does not come through as strongly in Boston as it would later in the Carnegie library program, there is an extent to which the BPL was also meant to be an engine of economic development on a broader social level.³⁶ For example, arguing in favor of the need for a public library in 1852, Everett asked:

Where is the young engineer, machinist, architect, chemist, engraver, painter, or student in any of the professions or of any of the exact sciences, or of any branch of natural history, or of moral or intellectual philosophy, – to get access to the books which are absolutely necessary to enable him to pursue his inquiries to any advantage? (reprinted in Winthrop, et al. 1855)

Building upon these rhetorical questions at the cornerstone ceremony at Boylston Street, Robert C. Winthrop³⁷ contends that:

Here, especially, will be collected without delay, whatever may throw light on the great practical arts which have characterized our age and country, and whatever may assist our ingenious mechanics and inventors, – second to none throughout the world, – in their attempts still further to simplify the magic processes, and to perfect the marvellous implements and engines, by which difficulties and distances may be annihilated (Winthrop, et al. 1855).

Essentially, by providing further means for self-education, particularly among practitioners of various useful trades, the library was meant to help Boston “promote its prosperity” (Seaver 1852) just as the public school system was seen to have done already. This motivation underlies a continuous thread through the development of the institution, as practical books became both a collection priority and a focus of several donations, including the British Government's gift of a complete set of the Publications of the Commissioners of the Royal Patent Office and the Bowditch donation of funds and materials to establish a medical collection (*Proceedings of the Trustees* 1852-1858, December 6, 1853, 54). Making these books available and useful to the relevant practitioners – for example, by keeping the library open into the evenings and on weekends and making the current periodicals of the day openly available in the Reading Room ("Report of the Examining Committee" 1864) – formed a key organizing principle of library structures and policies.

2.4.3 Public Demand

The question of the extent to which public demand played into the motivations for founding the Boston Public Library is an interesting one. The existence of such demand is

³⁶ This trope echoes even up to discussions of the value of public libraries in the present day – if anything, it's only gotten louder.

³⁷ President of the Board of Commissioners for the erection of the Library Building.

cited as a rationale for the library at various points in the documentation, but at those points it is always presented in the passive voice, with no supporting evidence. For example, the Boston Common Council (a branch of the City Council) describes the library as “an institution, the serious want of which is generally acknowledged” (1852b) – and the first Trustees’ report suggests:

a free public library is not only seen to be demanded by the wants of the city at this time, but also seen to be the next natural step to be taken for the intellectual advancement of this whole community and for which this whole community is peculiarly fitted and prepared ("Report of the Trustees of the Public Library of the City of Boston" 1852).

This seems a very top-down view. Perhaps more revealing than these very early documents, however, is way in which the concept of public demand is framed within the following explanation of the library’s success, written thirteen years later:

The cause of this success seems obvious. The Library had been for some time unconsciously called for, if such an expression may be allowed, by our community. It was, as Mr. Everett has said, in one of the previous Reports, simply the complement of our Common School System, and, as such, was needed, as its crowning grace in this city ("Report of the Examining Committee" 1865).

The fact that even in retrospect, the demand is depicted as hypothetical (or “unconscious”) tends to make one wonder whether the *actual* public demand for the institution was ever gauged in any way prior to the establishment of the library – or indeed, whether doing so was ever even considered. The leadership of the institution – from trustees to donors to city councilmen – certainly *assumed* that public demand existed. And they were proven correct after the institution opened. But there is little evidence that a serious effort was made to establish the existence of demand prior to setting forth on building the institution.

To be clear, the apparent lack of assessment of demand before the fact does not necessarily impugn the motives of the BPL’s early leaders. Rather, it points to the historical importance of the first iteration of the library – in those two tiny rooms on Mason Street from 1854-58 – as a prototype or proof of concept. As the Examining Committee would recall in 1863, before the Mason Street prototype illustrated the demand – and Bates’s generosity illustrated the support – for the library concept, “The project was regarded by many whose judgment and influence could neither be wisely nor safely overlooked, as an experiment promising little real or lasting good to the city” ("Report of the Examining Committee" 1863). It seems that the library leadership represented a contingent of Bostonians who believed in the existence of public demand, while other contingents did not; the early prototype at Mason Street tended to prove the more optimistic hypothesis.

3. Definitions

The foregoing section described the major rationales for the Boston Public Library expressed by the individuals and groups with most influence over the early years of the institution; this section will delve into the question of what exactly those individuals and groups thought – or hoped – they were doing, especially in terms of user base and collections. The first part, below, will explore how many and what kinds of people the leadership thought would be using the library, while the second enumerates the sorts of priorities and decisions that helped to shape the size and scope of the library's early collections.

3.1 Users

The discussion of users below will fall into two parts: first, an examination of the types of users the library leadership hoped or expected would use the BPL in the early years – that is, when they said “everyone,” which groups they actually seem to have *meant* – and second, a brief analysis of the actual scope of library usage, as reflected by the number of user registrations and the daily average borrowing statistics over the period from 1854-1865.

3.1.1 Intended Composition

The Boston Public Library trumpets its inclusiveness across the top of its front door, even to this day: Free to All. And indeed, in its formative years, the idea that the library should be open to *everyone* was an oft-repeated theme. The 1852 Report of the Trustees, for example, states that the library's collections:

should be kept in a Reading Room accessible to everybody; open as many hours of the day as possible, and always in the evening; and in which all the books on the shelves of every part of the Library should be furnished for perusal or for consultation to all who may ask for them, except to such persons as may, from their disorderly conduct or unseemly condition, interfere with the occupations and comfort of others who may be in the room ("Report of the Trustees of the Public Library of the City of Boston" 1852).



Figure 13: Entrance to Boston Public Library, Copley Square. Photo by author.

And later, the 1858 Examining Committee contends that the library “will the more fully accomplish its objects as its benefits are the more widely diffused – free to all who wish to enter its portals; free to use, but not free to abuse” ("Report of the Examining Committee" 1858). Many more such rhetorical exemplars exist. Of course, no physical library can be truly universal. Concerns such as space,

deterioration of materials, and limited numbers of copies – among many others – all serve to constrain the effective scope of the user base. And beyond these practical considerations, the image of the user in the minds of the leadership – the types of patrons those responsible for designing the institution had in mind while conceiving their designs – can have a real impact on the types of user best served by the institution: who feels welcome in the building, who can most easily navigate the catalogue, who finds the most materials of interest to them, and so on. Thus, despite the explicit desire for inclusiveness expressed by early BPL leaders, it is both possible and informative to tease out the specific user characteristics they had in mind as they gave the institution its initial form. In particular, this section will examine five types of user discussed by the BPL's early leadership: (1) Bostonians; (2) the needy, yet deserving; (3) both casual readers and serious researchers, (4) practitioners of the useful arts, and (5) women and immigrants.

3.1.1.1 *Everyone... in Boston*

The first boundary placed around the user base for the BPL in both rhetoric and rulemaking was the limitation of audience scope to *Bostonians* – that is, Boston taxpayers as well as certain other sorts of local people (on which more in a moment). Though the BPL was certainly the largest and likely the most inclusive public library in America (or possibly anywhere)³⁸ at the time of its creation, it was nonetheless designed primarily to serve the local community of Boston, and not the entire world. References to “all our citizens” (“Report of the Committee to Examine the Library” 1853), “the citizens at large” (Winthrop, et al. 1855), and “the greatest possible number of our fellow-citizens” (“Seventh Annual Report of the Trustees of the City Library” 1859) pervade the early internal and public documents of the institution, and reflect the strong focus on serving this local constituency exhibited by the early administrators and advocates of the institution. Of course, one should not put too fine a point on this limitation; after all, given the technologies of the era – books, shelves, circulation desks – it would have been much more difficult, simply as a practical matter, to extend the user base of the institution much further. In order to use the Boston Public Library in that era, one had to be physically present in the library building, at least from time to time. And politically, the focus on Bostonians also made a great deal of sense: when one is taxing a local community for the provision of a resource, it makes sense that it should be a priority to offer that local community first dibs on its use. Furthermore, focusing on the local population did not necessarily limit the scope of the institution a great deal, because its leaders had such enormous faith in the capacities of the local citizenry. At the laying of the cornerstone for the first dedicated BPL building on Boylston Street, the President of the Board of Commissioners for the erection of the Library Building averred that “No books,

³⁸ The 1866 BPL Examining Committee Report, among other sources, suggests as much: “The Committee suppose that no library in the world, of equal size, is so trustingly offered to a large population as this one is”(“Report of the Examining Committee” 1866)(“Report of the Examining Committee” 1866)(“Report of the Examining Committee” 1866)(“Report of the Examining Committee” 1866)(“Report of the Examining Committee” 1866).”

certainly, will ever be rejected in this land of universal education and intelligence, as being beyond the comprehension or capacity of the people. That comprehension will be subjected to no narrow gauge, nor that capacity measured by any reduced or stinted standard" (Winthrop, et al. 1855). Though the collection, like the user base, would not be truly comprehensive, it would not be out of condescension; the leadership clearly had faith that if a concept could be understood by anyone in the world, it could be understood by Bostonians.

3.1.1.2 *Those in need...but only the deserving*

As noted earlier in the discussion of egalitarianism and uplift, one of the founding principles of the Boston Public Library – one insisted firmly upon by both George Ticknor and Joshua Bates, respectively – was that it should strive first to serve those who could not afford to obtain books through other means. This theme is evident throughout the 1852 Trustees' report, but comes through most clearly in the sections attributed to Ticknor. At its plainest, the report decisively states:

Above all, while the rightful claims of no class, – however highly educated already, – should be overlooked, the first regard should be shown, as in the case of our Free Schools, to the wants of those, who can, in no other way supply themselves with the interesting and healthy reading necessary for their farther education ("Report of the Trustees of the Public Library of the City of Boston" 1852, 15 (marked 17)).

Bates, in turn – having grown up poor himself – also emphasized the value of the library to those in need, not only because it would provide intellectual stimulation, but because it would serve as a wholesome and welcoming third-place:

My experience convinces me that there are a large number of young men, who make a decent appearance, but living in boarding houses or with poor parents cannot afford to have fire in their rooms. Such persons in past times, having no place of resort have often loitered about the streets in evenings and got into bad company, which would have been avoided, had such a library as is now proposed been in existence. The moral and intellectual improvement such a library would produce is incalculable (Trustees of the Public Library 1865, letter from Bates to Thomas W. Ward, November 5, 1852).

Though the ultimate goal was to build a library that would serve all of Boston – well-to-do and working class alike – the initial priority was to provide for the needs of those less able to provide for themselves. This would have implications for collection development, building design, and cataloging priorities, as will be discussed later in this chapter.

Still, there was a tension in the rhetoric surrounding these needy individuals the library was intended to serve. In particular, it typically included several caveats specifying that only a particular *sort* of person-in-need would truly be welcome in the library. That is, those who are "virtuous and industrious" (Bates 1852) and "desire earnestly to improve their

minds" (Winthrop, et al. 1855, letter from Everett, August 12, 1850), and who do not "from their disorderly conduct or unseemly condition, interfere with the occupations and comfort of others who may be in the room" ("Report of the Trustees of the Public Library of the City of Boston" 1852, 19). The sense that the library truly only welcomed "the good kind" of needy people is further substantiated by the rules governing access to the institution, which specify that registered users of the Reading Room must be "of respectable character, and of such orderly conduct and condition as not to interfere with the occupations and comfort of others, who may be in the room at the same time," while borrowers had to meet more stringent character-proving requirements such as being known or vouched for personally to the Librarian. Furthermore, if an individual could not be adequately vouched for, they would have to deposit the value of the book in order to borrow it – not a small requirement given the relatively high cost of books at the time (Board of Trustees 1853). It thus seems reasonable to suggest that borrowing was only available in a limited way to those in need, given that one suspects members of the poor and working classes might have had both (a) greater trouble getting someone with validity to the library administration to vouch for their character and (b) insufficient means to place a deposit on a book.

3.1.1.3 Both casual readers and serious researchers

From the beginning, the BPL was intended to a hybrid of a casual-use public library and a serious public research library. In the earlier years, tied to the emphasis on "those in need," the focus in budgeting and policymaking was decidedly on the casual-use side, with research works coming in to the library only through donation as the entire collections budget was expended on works – often in many copies – of popular interest for lending (e.g., "Second Annual Report of the Trustees of the City Library" 1854, 7). Further, the library was divided into a Lower Hall, which housed the lending library and popular works, and an Upper Hall, which housed the reference collection, the circulating research collection, and valuable or rare materials which would not circulate; when time came to catalog these collections, the Lower Hall was the clear priority, and was completely cataloged within months of the 1858 opening of the Boylston Street building, while the Upper Hall catalogue did not reach a similar level of completion until 3 years later (Whitehill 1956, 58-59). Much of this priority-setting was attributed to Ticknor, who insisted that the library focus on popular use first, before delving into the service of serious researchers who were often already well-served elsewhere in the region (e.g. Harvard, the Boston Athenaeum, etc.). As his biographer notes, Ticknor's

desire to put culture within the reach of those who are least apt to seek it and least able to acquire it, and his belief that they could be trusted to use carefully what was bestowed generously, this desire and this belief inspired his action for the Library for the first six or eight years of its development; but when the principles he thus contended for were vindicated by experience, and put beyond danger, he turned to work for the more scholarly and studious class, of which he

himself was a member (Hillard, et al. 1876, 319).

As went Ticknor, so went the library: by 1857, the leadership had begun to focus more on establishing the library as a destination for researchers. As a later Examining Committee report explains,

Up to 1856 the system of purchases had looked to supplying the most popular wants. The collection, which had then grown to near 30,000 volumes, was deemed large enough to satisfy the most reasonable demands of a general kind; and it began to be felt that there were particular classes of our citizens, apart from the general body, whose wants deserved recognition. So about that time we find, that books in the foreign tongues began to be added, and the higher departments of literature more fully developed ("Report of the Examining Committee" 1867).

Still, because the library leadership had such great faith in the capacities of the local population, they also saw the potential for users of popular materials to *become* serious researchers – indeed, this principle of “taste elevation” was a major strategic point around which the library was organized. The idea was that the library’s patrons would begin by reading popular works of somewhat lesser substance, and would gradually progress to develop a taste for more serious reading and research. Thus, although serious researchers and casual readers were frequently depicted as separate sets of individuals, there was a definite hope among the leadership that library users would progress from one into the other, like caterpillars turning into butterflies; that readers would start off with the popular literature of the day, and progress (perhaps through biography or history) into the reading of more serious, educational materials (e.g., "Report of the Trustees of the Public Library of the City of Boston" 1852, "Second Annual Report of the Trustees of the City Library" 1854, Capen 1900?). The dual focus of the institution remains in place to this day; however, to my knowledge, the transformative hypothesis above has never truly been proven.

3.1.1.4 Practitioners of the useful arts

Somewhere between casual readers and serious researchers, moreover, existed a third class of favored library users: practitioners seeking information related to their fields of practice. As noted earlier, one motivation for establishing the public library was to make it into an engine of social and economic progress. As such, a strong emphasis was placed on serving practitioners of a wide range of useful functions in society: “the young engineer, machinist, architect, chemist, engraver, painter, or student in any of the professions or of any of the exact sciences, or of any branch of natural history, or of moral or intellectual philosophy” ought to be able to freely access “the books which are absolutely necessary to enable him to pursue his inquiries to any advantage” (Winthrop, et al. 1855, letter from Everett, June 28, 1852). The provision of periodicals in the Reading Room was seen as being especially important to fulfilling the needs of this user group. As the 1864 Examining Committee would approvingly note:

The Reading Room, which is one of the most attractive departments, is appreciated, affording to the poorest artisan the opportunity, which would otherwise be confined to the rich, to consult the weekly and monthly record, foreign and domestic, of the progress making in arts, sciences, and manufactures; often enabling an humble mechanic to obtain information of some new discovery in mechanics, or some new principle applicable to the industrial pursuit in which he is engaged, before it has attracted the attention of his employer. The current intelligence of the day flows into the minds of the people, giving new impulses to the brain and quickening the best energies of men ; and it is, therefore, on this account, that we recommend that the Reading Room should be made the receptacle of all magazines, at home and abroad, which may prove sources of literary entertainment, or, what is still more valuable, contribute to the wellbeing of the laboring classes, by bringing constantly before them the news which has a direct bearing upon the avocations in which they are engaged ("Report of the Examining Committee" 1864).

This was seen as an extremely valuable function of the library, as existing information repositories in the area – especially the Boston Athenaeum – were not sufficiently open and accessible to fill this role, despite the clear pragmatic benefits of a well-informed class of, say, physicians.³⁹

3.1.1.5 *Women and immigrants*

Though women and immigrants are mentioned in fewer places than the above elements of intended audience, it is nonetheless interesting to note that the library's leadership expressed the intention from the earliest days of planning to extend the benefits of the library to these two groups on broad terms. Women, in particular, were offered access to the library on equal footing with men, and formed a significant part of the library staff as well. In fact, initially there was to be a women's reading room in the Boylston Street building, but it was quickly discovered that mixing the sexes in a single room produced no ill effects, so the designated room was turned to other uses (e.g., "Thirteenth Annual Report of the Trustees of the City Library" 1865, "Report of the Examining Committee" 1865). Shortly after the opening of the Boylston Street building, the Trustees would declare that

[they] regard it as one of the most pleasing and hopeful features of the establishment, that its advantages are equally open to both sexes. Every thing in their power has been done by the Trustees to encourage the resort of both sexes to the Institution, whether for the loan or consultation of books. More than half the assistants employed in the Library are females, and a separate reading room for ladies will be opened, if it is found desirable. Thus far, however, neither in Mason street nor in the new building, has any inconvenience resulted from the use of a single reading room. By nothing has the administration of the Library been more signalized than the order and decorum which have uniformly prevailed among those frequenting it ("Sixth Annual Report of the Trustees of

³⁹ This group was specifically called out as an important one to support in the 1866 Report of the Examining Committee.

the City Library" 1858).

Though the early documentation of the library does exhibit some of the general nineteenth-century tendency to use the word "men" in place of "people," the phrase "both sexes" is used with sufficient frequency to surmise that women were, in fact, very much an intended part of the library's user base (e.g., "Report of the Trustees of the Public Library of the City of Boston" 1852, Winthrop, et al. 1855, letter from Everett, June 28, 1852, "Fifth Annual Report of the Trustees of the City Library" 1857, "Sixth Annual Report of the Trustees of the City Library" 1858, "Report of the Examining Committee" 1866). Still, it is intriguing to note the difference in gender representation between the drawing and the photograph of the Lower Hall Reading Room reproduced in Figure 14. Perhaps the day the photo was taken was simply an off day, or the wrong time – or perhaps women were not as prone to use the Reading Room as the Trustees might have hoped?

The tide of immigration in the middle of the nineteenth century also had an impact on the leadership's visions of what their user base might be. This rising tide is noted in the documents as early as 1852, and by 1857, it is noted that:

They have also added considerably to the number of volumes in the principal languages of continental Europe. There is in Boston a large population which speaks some one of those languages as their mother tongue. All will feel that it is for the public benefit, that this class of our citizens should have the means of improving their minds in common with the rest of the public ("Fifth Annual Report of the Trustees of the City Library" 1857).

Though immigrants were by no means primary in the minds of library leadership, this nod to the need to serve them – or perhaps rather, to *educate* and/or *improve* them – interestingly



Figure 14: Two depictions of the Lower Hall Reading Room at Boylston Street. Images courtesy of the Boston Public Library, on Flickr.

presages the Americanization movement that would begin to sweep through public libraries in the 1890s (e.g., Harris 1975a, 13-14).



Throughout these visions of the user, there runs a perceptible thread of paternalism, and in some cases also elitism. Michael Harris, in fact, goes so far as to suggest (following an unrelenting takedown of Ticknor) that Ticknor's goals for the library, reflecting his "authoritarianism and elitism," actually boil down to two points:

(1) to educate the masses so that they would follow the 'best men' and not demagogues; to 'stabilize the republic and to keep America from being a second Carthage;' and (2) to provide access to the world's best books for that elite minority who would someday become leaders of the political, intellectual, and moral affairs of the nation (Harris 1975a, 8).

And there is an element of truth to this, reflected particularly in the discussion of casual readers and serious researchers above. Still, though influential, Ticknor was not the only relevant decision-maker for the library – and even were he so, some of Harris's critiques strike me as exaggerations. Rather, there seem to have been quite a number of considerations present in the minds of the leadership when pondering the institution's user base – yes, whether or not the individual was well-behaved and deserving, and yes, training up casual readers into better citizens and possibly serious researchers (in the image of men like Ticknor and Everett, both Harvard faculty), but also local citizenship, degree of need, social requirements for specialized practitioners, and a desire to include – and/or perhaps indoctrinate – more marginalized portions of the population like women and immigrants.

As it happens, successive revisions of the library rules during the period included in this study suggest that the actually-allowed user base of the library grew broader, little by little, over time. Initially, borrowing privileges were limited to (1) adult inhabitants of Boston of good character (they had to be vouched for or known to library officials), (2) local clergymen and teachers, (3) particularly good upper-level students from the Boston public schools, and (4) other inhabitants of Boston (presumably those whose character could not be verified) willing to deposit the value of the volume asked for (1853). Over the years, this list was expanded, and its conditions revised. By 1858, several additions had been made, including any individual donating \$100 or more to the library, all students at the Boston Normal School and Girls' High, city government officials, individuals not residing in Boston but paying Boston taxes, and individuals willing to make a co-op style deposit that would grant them a voucher for borrowing – initially, \$2, with the understanding that taking out books of greater value than that amount would require an additional deposit (*Proceedings of the Trustees 1852-1858*, September 13, 1858, Board of Trustees 1858).

3.1.2 Quantity

About the quantity of users expected and later experienced by the library, there is less to report, though a few basic statistics may be helpful for orientation as to scale. There is little indication of the actual numerical expectations for the popularity of the library; indeed, as noted previously, there was a non-insignificant segment of the population who were not certain that the library would be useful at all. However, the actual usage of the library is extremely well documented, and can be shown in a few different ways.⁴⁰ First, growth in user registrations over time. Figure 15, below, shows the cumulative number of individuals who had registered to use the library; in the years 1854-1857, this reflects the number who signed up to use the reading room, which at that time was separate from the registration process for borrowing (the number of registrations for borrowing was consistently smaller, and it is reasonable to surmise that it represented a subset of those who registered to use the reading room). In 1858, when the Boylston Street building opened, a new, integrated system of registrations was begun, and the thousands of existing registrants made to re-register. This accounts for the dip in the center of the graph. Overall, the growth in usage is almost linear, adding roughly 4,000-6,000 names to the rolls each year. For scale, the population of Boston in 1860 was 177,840; by 1870 it had grown to 250,526.⁴¹ Assuming for a moment a linear rate of growth between those years, the number of library users expressed as a percentage of the population would have grown from about 10% in 1860 to about 19% just five years later. For a new institution that still only had one (extremely crowded) branch, this seems a quite respectable level of early-adopter buy-in.⁴²

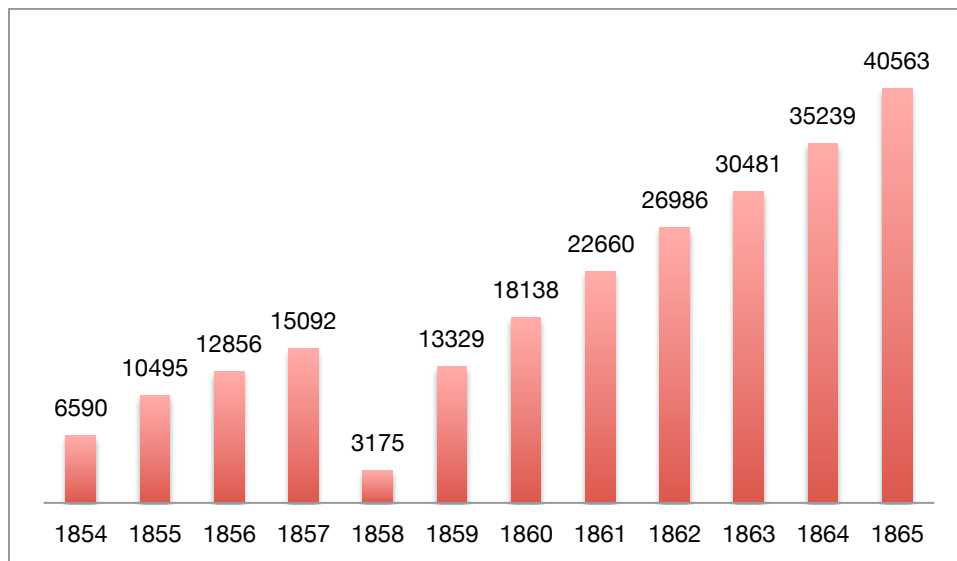


Figure 15: Cumulative User Registrations at the Boston Public Library, 1854-1865

⁴⁰ Data for the charts was assembled from the Trustees', Superintendent's, and Examining Committee reports for the relevant years.

⁴¹ http://www.census.gov/history/www/through_the_decades/fast_facts

⁴² By way of comparison, in 2008, the ALA found that 68% of Americans had library cards (Peterson 2008).

In terms of actual lending, the usage of the library grew quite consistently year over year for the first decade, with 1859 likely representing a temporary aberration related to excitement over the recently-opened new building (Figure 16). In 1867, the (unprecedentedly detailed) Report of the Examining Committee would compare Boston's lending against other libraries of similar size and stature, and conclude that given the city's lack of branch libraries and relatively small population, the BPL's lending seemed at least adequate to expectations ("Report of the Examining Committee" 1867, 39-43).

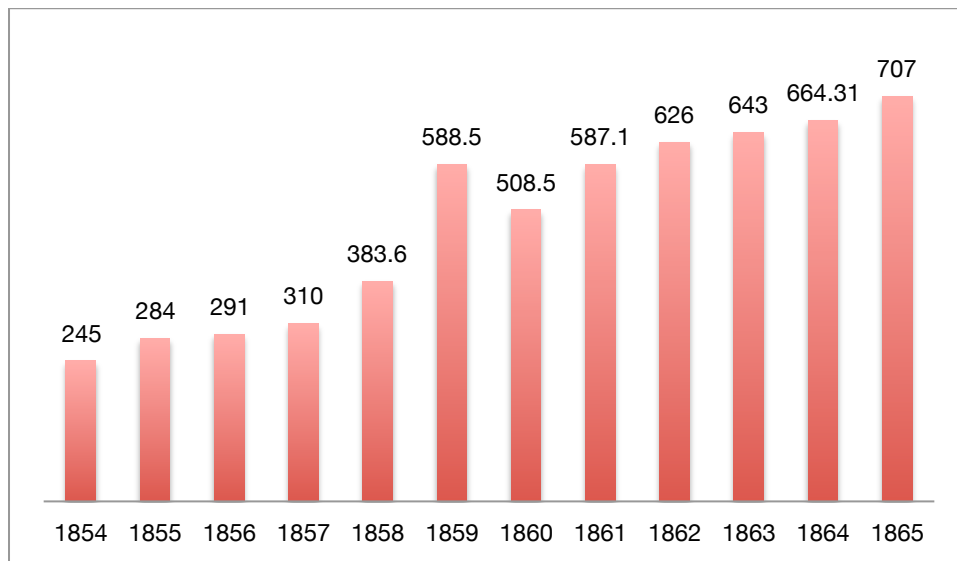


Figure 16: BPL Daily Average Lending, 1854-1865
(Figure for 1854 is an estimate, based on total circulation and approximate number of open days)

The extent of use of the library had significant implications for the structures and processes developed to meet that demand. Though those are topics to detail in later sections, it will be useful to note for now that the popularity of the library created continual space problems throughout the early history of the institution. As noted in the historical synopsis, it was clear before the library even opened at Mason Street that the two rooms there would be far too small; in 1856, when the rooms had been in operation for two years, the Examining Committee would assess the situation as follows:

Indeed, the inadequacy of accommodations, and the inconveniences attending the operations of the library, so fully set forth in the reports of our predecessors, have seriously increased — in view of which it is highly creditable to the librarian and his assistants, that the varied, extensive, and constantly augmenting business of the library has been carried on with promptness and efficiency. But the evils and embarrassments which have hitherto surrounded the institution in regard to locality and narrowness of its borders, are, thanks to the judicious and munificent provision of the municipal government, soon to pass away. Upon the completion of the edifice designed for its use, its operations will be conducted on an enlarged scale, and its privileges be enjoyed by great numbers of citizens, to whom, thus far, they have been practically denied. The institution has

accomplished much good under many difficulties; from small beginnings it has become, within a brief space of time, one of the principal libraries in the land; and, if continued to be wisely managed, it is safe to predict for it a long and prosperous career, extending its blessings to remote generations of our descendants ("Report of the Examining Committee" 1856).

Though the library would move into a dedicated building in 1858, that too would quickly prove too small to contain either the burgeoning collections or the throngs of readers who flocked to use them (much less the staff space necessary for things like cataloging and binding) (Whitehill 1956, 79-80). This was partially a matter of poor building design; however, the extent to which usage of the library exceeded the leadership's expectations clearly played a role as well.

3.2 Collections

The above discussion provides the outlines of the BPL leadership's definition of "everyone;" the section that follows will explore their ideas about "everything." In particular, reversing the order of the previous section, this section will first offer a brief discussion of the size of the BPL collections, as they grew from 1852-1865, and then explore the way in which priorities were set for building those collections.

3.2.1 Size

From 1852 to 1865, the Boston Public Library grew from a modest collection of about 4000 donated volumes to become one of the largest and most valuable collections in the world. By 1867, in fact, the library had become the second-largest library in the United States, surpassed only by the Library of Congress ("The Experiment of Free Libraries" 1867). This steady - and occasionally explosive - growth is illustrated in Figure 17.

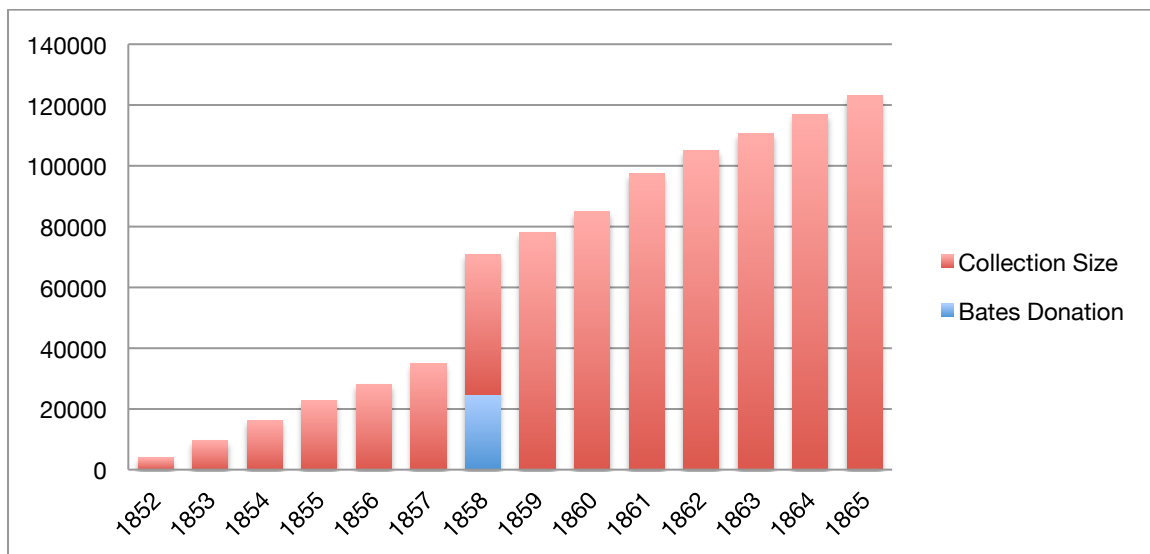


Figure 17: Growth in BPL Collection Size, 1852-65
(Data from Annual Reports of the Trustees, Librarian/Superintendent, and Examining Committees)

At least two things bear noting about this chart. The first is the relatively tiny height of the first bar, for 1852, which reflects the extent to which the library was first established as an experiment, or proof of concept. In their report for that year, laying out the plan of the library, the Trustees advised that:

In the establishment of such a library, a beginning should be made, we think, without any sharply defined or settled plan, so as to be governed by circumstances as they may arise. The commencement should be made, of preference, in a very unpretending manner; erecting no new building and making no show; but spending such moneys as may be appropriated for the purpose, chiefly on books that are known to be really wanted, rather than on such as will make an imposing, a scientific or a learned collection; trusting, however, most confidently, that such a library, in the long run, will contain all that anybody can reasonably ask of it ("Report of the Trustees of the Public Library of the City of Boston" 1852, 19).

In order to make the proposed library palatable to a reluctant City Council, the Trustees recommended starting it off at a very conservative scale: no dedicated building, no unnecessary ostentation – just a small appropriation from the Council for the purchase of a modest set of popular books, with the hope that more scholarly volumes would follow later on, after the library had proven its worth to the masses.

The second notable item in the chart above relates to the second Bates donation, offered in 1855 and incorporated into the collection by 1858 (in blue in Figure 17). This donation added 24,618 volumes to the collection, which in the previous year had totaled 34,896. Thus, on its own, it increased the size of the collection by 80% relative to the previous year; in combination with other acquisitions (both donations and purchases), it more than doubled the collection between 1857 and 1858. This enormous donation essentially turbo-charged the library's growth, and thus its status as a collection of significant scale. Ironically, however, it would also contribute to the rapid obsolescence of the Boylston Street building, as it carried out the donor's wish to "provide books to fill the new rooms" all too well (Trustees of the Public Library 1865, letter from Bates to Everett, June 29, 1855) – leaving insufficient space for the subsequent growth of the library in terms of both collections and use.

3.2.2 Composition

The leadership's pride in the size of the collection is reflected throughout the reports of the Trustees, Examining Committees, and Librarians/Superintendents. The sense that a bigger collection would be a better collection, and that accordingly, no donation or appropriation was likely to be turned away, pervades these early documents. Though some exceptions to

this generality were made in the case of books deemed “immoral”⁴³ (e.g., *Proceedings of the Trustees* 1852-1858, May-June 1858), these appear to have been few and far between in the early years. Still, the leadership also had very particular ideas about what sorts of things this collection *should* contain, and what the library’s priorities for acquisition should be. For example, though both popular and scholarly books were to be included in the BPL’s collection from the start, the proportion and priority given to each type shifted over time, in keeping with the philosophy of “taste elevation” that guided much of the thinking on collections by the trustees and librarians of the institution at that time (e.g., Williams 1988, 5-7). Thus, early on, the budgetary priority for the library – that is, what the library’s own funds were expended to purchase – was popular books, in multiple copies. The idea was that the populace would be attracted to the library to read these books, but would come to desire better books over time. The popular books were thus a gateway to “higher” forms of reading. This also helps to explain why the early purchasing guidelines for the BPL demanded that the books:

must be cheap: selecting from the catalogue for purchase only those books that can be bought by auction or otherwise at trade prices or below. But other things being equal, we wish you to purchase of preference those books that are of a popular and practical character, such as books of History, Biography, Voyages, Travels, and books relating to the arts and the business and duties of everyday life; although we shall still be glad to have you buy any book whatever on the list, when a copy of it turns up particularly cheap (*Proceedings of the Trustees* 1852-1858, July 5, 1853).

In this period, the vast majority of rare, valuable, and/or particularly scholarly books acquired by the library were received as donations – especially from the library’s largest benefactor, Joshua Bates, but also from many other smaller donors (e.g., *Proceedings of the Trustees* 1852-1858, October 19, 1858). Via the concept of taste elevation, the prioritization of works with mass appeal tied in with the “uplift” motivation discussed earlier. In 1854, the Trustees explained how they expected this process to work, and how their initial expenditures on popular works fit into the plan:

It is true that a considerable number of the new books added to the library the present year are duplicate sets of popular new publications, which have been purchased in pursuance of the principle on which the circulating department is founded, viz : that of supplying the public, as far as it could be reasonably done, with those books for which there is an urgent present demand. It is expected that for a certain class of books this demand will be temporary, and that in due time – sooner for some and later for others, – all but a single copy, or a very few copies may be dispensed with. In the meantime, however, the books will have performed their office; the taste and habit of reading will have been cultivated; and if the expectations of the trustees are realized, the standard of the reading

⁴³ And so the discussion about public libraries shall ever go...

community will have been steadily raised, and the class of new books called for will become more and more one of permanent value ("Second Annual Report of the Trustees of the City Library" 1854).⁴⁴

The focus on popular works in the early years was further emphasized by the priority given to cataloging the lower hall of the library (which contained the circulating collection) before the upper hall (which contained the reference and scholarly collections) as the library moved into the Boylston Street building. This, like the prioritization of purchasing popular works, was due to Ticknor's insistence (Whitehill 1956, 57, Tyack 1967, 211). As one of Ticknor's biographers notes, the complete catalog, containing both upper and lower halls "was a more impressive tome, but not so useful to the 'less favored classes'" whom Ticknor saw it as the library's special purpose to serve; upper- and middle-class scholars, after all, already had excellent resources available to them at the Athenaeum and elsewhere (Tyack 1967, 211). Ticknor remained devoted to the idea that the library could uplift and educate these classes in particular, and his influence over the library's administration remained undiminished nearly until his death; the library's initial focus on both collecting and cataloging the more popularly accessible works is in no small part attributable to that influence.

Still, there was a difference in the leadership's minds between "popular" and "light." And whereas it was a primary objective of the library to provide the former, the latter could not have been frowned upon more forcefully. The distinction between the two is highlighted in the Trustees' Fifth Annual Report:

The Trustees, in the purchase of books, have adhered to the principles which have hitherto guided them, of keeping the library supplied with the current literature and fresh reading of the day. They have aimed to add useful books to the library, rather than what is called "light reading." The best interests of the institution require that it should not be regarded as a depository of books of the latter description. They are so cheap that they can be otherwise obtained by almost every one who wishes to read them; they occupy space on the shelves better filled by better books; and they increase the resort of persons to the library whose wants might be easily supplied in other quarters, leaving the Librarian and his assistants to devote their attention to more earnest and thoughtful readers. The Trustees are persuaded that it was not the design of the judicious and public-spirited citizens who, as members of the City Council in years past, or at the present time, have liberally appropriated the public funds to the foundation and support of the library, to have it become the means of

⁴⁴ And although the historical record as a whole is more ambivalent about the efficacy of the taste elevation hypothesis, the Trustees were already claiming success as of their third annual report, stating that: "A higher class of books appears to be gradually called for, showing that the taste for reading improves with the increased means for its indulgence. Enquiries are constantly made for works of scientific and literary utility; and the Trustees have had the gratification of supplying this demand in cases where it could be satisfied from no other quarter in this City" ("Third Annual Report of the Trustees of the City Library" 1855).

gratuitously supplying to a class of idle readers, the unprofitable, not to say pernicious trash, which is daily pouring from the press ("Fifth Annual Report of the Trustees of the City Library" 1857, 4-5).

This distinction reveals a tension in the Trustees' thinking, not just about the desired contents of the library's collections, but also about their desired user base. On the one hand, they clearly wished the library to appeal to the popular taste, and for it to serve those who, in Ticknor's words, were "less able than they would gladly be to procure pleasant and profitable reading for themselves and for their families" ("Eighth Annual Report of the Trustees of the City Library" 1860). On the other hand, however, this desire to serve those in need was continually accompanied by caveats regarding the utility and constructiveness of the reading to be offered. And those caveats, especially when paired with statements of policy like the one above, tend to reveal at least a lack of faith in the library's users, or at worst, a lack of respect for their autonomy as people who deserve to make their own decisions about what they read. There is a definite sense that although the library leadership highly valued the institution's role in serving those with lesser economic means, those leaders also saw at least some of those individuals – particularly some of the more recent groups of immigrants (e.g., Tyack 1967, 222, Harris 1975a, 6) – as being just generally *lesser*.⁴⁵ This type of user, in turn, stood in contrast to scholarly users, more like the Trustees themselves, who could be trusted to use the library for its intended educational purpose without as much paternalistic gatekeeping on the part of the leadership.⁴⁶



By the late 1850s, however, the Trustees' sights had turned toward procuring more scholarly works, to supplement the popular collection and the continuing stream of donations. For example, before Ticknor departed on his purchasing trip to Europe to expend the second Bates donation in 1856, he collected from several eminent scholars "lists of works

Letter written by Ticknor to Prince Johann of Saxony in 1854, in which he suggests that "the vast immigration of late years from Europe, ... though the respectability of the mass of persons who, in general, were fitted to understand the principles of universal suffrage" (Georg and Daenell 1920, 95).

⁴⁶ It is also worth noting, as a last statement on this point, that the claim, above, that light reading was "so cheap that [it could] be otherwise obtained by almost every one who wishes to read [it]" seems to run directly counter to the institution's aim to serve those most in need: that is, those who actually could *not* afford *any* books, no matter how cheap. The cost of the books, however, was likely a secondary consideration in this case, simply offered as an additional rationale for excluding the books mainly on the basis of their content.

in various departments of knowledge, which they deemed most important for a collection of this kind" ("Fifth Annual Report of the Trustees of the City Library" 1857, Superintendent's report), which he then duly sought out all over Europe. And a later Examining Committee, looking back, notes that by 1856,

The donations to the Trust Funds, now accruing, in being expended for books of solid and permanent value, served to strengthen very materially the upper classification; while Mr. Bates' last munificent gift of books developed our weight in the same direction. The time was now come when it was very properly agreed that there was no department of learning, which some portion of the community was not interested in; and that every department should be cared for to meet such requirements. So the two distinct collections have been developed – the Lower Hall to meet the ordinary demands of the people, and the Bates⁴⁷ to serve the higher requirements of the studious classes, or of investigators in special matters – a scheme which your Committee cannot but think naturally evolved, and conducive to the satisfaction of every mental grade, and answering the requirements of all the intellectual demands of the community ("Report of the Examining Committee" 1867).

By the 1890s, during the period of "rule by committee," the library had tilted strongly enough toward the acquisition of rarities and scholarly collections – and away from popular works – that the Examining Committee admonished the Trustees to remember the mission of the institution, to diffuse information to the broadest possible swath of the population – and not just to serve the needs of small groups of scholars (Whitehill 1956, 123-24). Still, throughout the earlier period of interest in this dissertation, the institution seems to have maintained a relatively balanced dual focus between popular and scholarly books – though with an ever-strengthening resolve that both categories of books should be worthy and useful, and not mere ephemera or trash.

4. Implementation

Having discussed the motivations of the Boston Public Library's leadership in founding the library, as well as their conceptions of whom the library would serve and what it would contain, the next logical step will be to ask how those motivations and conceptions were put into practice. To some extent, this has been discussed already, as practice forms some of the most compelling evidence available for ascertaining concepts and rationales that might not make their way into explicit text; to paraphrase a couple of clichés, both actions and physical objects can often speak louder than words. Still, it will be useful to more directly examine some of the structures and processes put into place in the early BPL, while also assessing their connections to the motivations and definitions described above. The final section of this chapter will thus do just that, describing first some of the physical structures,

⁴⁷ The Upper Hall at Boylston Street, like the Reading Room at the current Copley Square Building, was named "Bates Hall," after the library's early benefactor.

and second some of the administrative processes, put in place to enable and/or regulate use of the library and its collections.

4.1 Structures

In this dissertation, I use the term “structure” to describe elements of the initiatives which have an independent existence outside of human action (e.g., once a building is constructed, it remains standing even if nobody is using it) and which affect the way users navigate the resources the initiatives create. For the Boston Public Library, the most influential structures in the early years were threefold: (1) the building at Boylston Street, which was the first attempt at designing an edifice explicitly for use as a central, urban public library in the United States; (2) the arrangement of the book stacks, and in particular their inaccessibility to the public; and (3) the library catalogs, of which there were more different types than one might expect, and in which BPL librarians were key innovators.

4.1.1 Building

The first dedicated home of the Boston Public Library on Boylston Street was razed four years after the removal of the library to its current quarters in Copley Square, in 1899 (Wadlin 1911, 66). As such, the building’s structure and character can only be ascertained from written descriptions, architectural plans, and a small number of late nineteenth-century photographs.⁴⁸ Still, from these

documents, it is possible to assess both the design of building itself and the connections of that design to the social motivations and assumptions outlined above.

Special Requirements for the Building

1. A library hall, capable of containing, at least, two hundred thousand volumes. The alcoves, or other divisions of this hall, are to be easily accessible from each other. All the divisions and shelves are to be so contrived as to follow strictly the decimal system⁴⁹ ; that is, each of its principal and lesser divisions is to be in series of tens. The shelves are to be fixed, and not movable.
2. A general reading room, with ample accommodations for at least one hundred and fifty readers at tables.
3. A special reading room, for ladies, with seats for at least fifty persons.
4. A room for the delivery of books to borrowers, which may be used as a conversation room, with accommodations sufficient for at least two hundred persons.
5. A library room, connected with the foregoing, in which not less than twenty thousand volumes of books most constantly demanded for circulation, may be arranged on the decimal system.
6. A Trustees' room, of moderate proportions.
7. A Librarian's room, etc.

⁴⁸ On the bright side, all of these materials are presumptively in the public domain in this country, and thus can be reproduced here. Many have also been digitized, and are available (a) on the BPL’s website, (b) via the Internet Archive, and/or (c) on Flickr.

⁴⁹ Despite the specification of a “decimal” system, this should not be confused with the later system devised by Melvil Dewey. The classification system of the early Boston Public Library was created by a trustee, Nathaniel B. Shurtleff, and referred to physical shelf position. As Whitehill explains, “the fixed shelves were so numbered that the figures in the place of hundreds denoted the alcove, those in the place of tens the ranges, and those in the place of units the shelves. Thus a book numbered 2236 would obviously be found, even by the dimmest intelligence, on the 6th shelf of the 3rd range of the 22nd alcove” (1956, 67).

The design for the Boylston Street building was determined through a competition among local architects, in

Figure 19: Special Requirements for the Building (Boston City Council 1858, 139)

which the participants were given two lists of requirements – one “general,” one “special.” The general list included several items that would have been of concern in the construction of any public building of the time – resistance to fire and sufficient lighting, ventilation, water, and heat – but also more library-oriented practicalities: avoiding dampness, so as to preserve the books, and allowing for centralized supervision and management, “so that the Librarian and his assistants can be readily informed of all comers and goers, and that the care of the library may be intrusted to as few officers as possible” (Boston City Council 1858, 139). The second, “special” list (Figure 19) spoke more directly to the perceived needs of a public library building – though unlike the general list, which spoke more to practical necessities, the special list looks more like the Trustees’ wish list: not what the library *must* contain, but what it *should*. This included shelving and reading room capacity (#1, #2, #3, and #5),⁵⁰ space for delivering books to users (#4), rooms for the Trustees and the Librarian (#6 and #7), and a division in storage space between the popular (#5) and less-frequently-used (#1) portions of the collection. One interesting point of note regarding both lists is their complete omission of any reference to aesthetics, which was likely intentional. Overall, the library leadership appears to have been a great deal more concerned with functionality than with appearances when it came to this building.⁵¹ As early as 1850, Everett suggested that the plan for the library should “[aim] at nothing but convenience and neatness,” adding parenthetically that “all attempts to go farther in architecture are almost sure to fail” (reproduced in Winthrop, et al. 1855). And at the ceremony for the laying of the cornerstone, the President of the committee on the building displayed a mix of pride and humility in declaring that the building

⁵⁰ It is worth noting that the requirement that the reading room seat at least 150 comes directly from one of Joshua Bates’s donation letters (Trustees of the Public Library 1865, copy of letter dated October 18, 1852). Though ultimately it would prove a gross underestimate, it is nonetheless indicative of the influence over the institution that went along with Bates’s large donations (an influence which, given its potential strength, he wielded extremely rarely).

⁵¹ The McKim Building at Copley Square would be another matter entirely; a later history of library architecture would call it “an Italian palace...with a library fitted into it, but with such loveliness of proportion and detail that its defects as a functional plan are overlooked” (Wheeler and Githens 1941).

may never vie, indeed, with the sumptuous libraries of the old world, in the magnitude or magnificence of their structure, or in the costliness and rarity of their contents. We have aimed at no imposing facades, or splendid colonnades. But it is confidently believed, that, by the skill of our ingenious architect, ... few buildings will be found to equal it in practical appropriateness and convenience (Winthrop, et al. 1855).

The “ingenious architect” in this case, as previously noted, was Charles K. Kirby, whose design was selected out of 24 entries. His plans, portions of which are reproduced here, remain in the collections of the Boston Public Library Fine Arts department, along with a handful of the rejected designs.

As the previous discussions of collections have indicated, the building Kirby designed had two levels open to the public.⁵² On the first floor (Figure 20), the entryway opened on a vestibule housing an impressive double staircase, which led up to the second story. Straight ahead, between the stairs, a doorway opened onto the Delivery Room (marked “Conversation Room” in the plans), where books could be requested. Popular books, as noted earlier, were kept in book alcoves on the lower level (marked “Library Room”). Through the book alcoves, the basement was accessible via the staircase pictured in the upper left, and the upper stacks via the spiral staircase in the upper right. A small room for the librarian⁵³ was also provided here. In the Delivery Room, there were several pew-type wooden benches on which to await requested materials; more seating was available at long tables in the main Reading Room, accessible through a set of doorways to the right (See also Figure 14, above). To the left, and back toward the entrance, a door opened onto a second reading room containing periodicals and reference materials (marked “Special Reading Room”), which

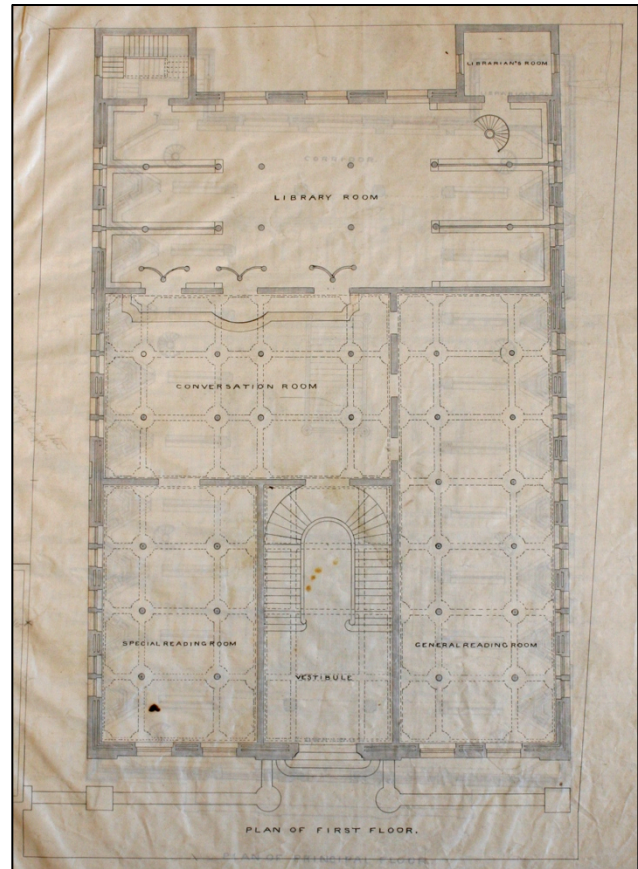


Figure 20: Plan of the First Floor of the Boylston Street Building, Charles K. Kirby, 1857. Courtesy of Boston Public Library, Fine Arts Department. Photo by author.

⁵² The basement was also used for library purposes – though a large part of that space was occupied by the furnace and the fuel with which to feed it.

⁵³ Or perhaps for the Trustees; a Trustees room is called for in the requirements, but there is no room in the plans labeled as such – only two rooms labeled as Librarians’ Rooms.

had been intended as a ladies' reading room, but which was never used as such (Kirby 1857, Whitehill 1956, 59).

The upper level (Figure 21) was entirely occupied by Bates Hall, a monumental, high-ceilinged room flanked on all sides by towering columns, in which oval-shaped tables for reading were provided in the open space surrounding the central staircase. Cabinets and iron railings separated this open public area from the tiered book alcoves that climbed three stories high on every wall, which were only accessible to librarians. At some point, a staff desk was added at one end of Bates Hall as well.⁵⁴

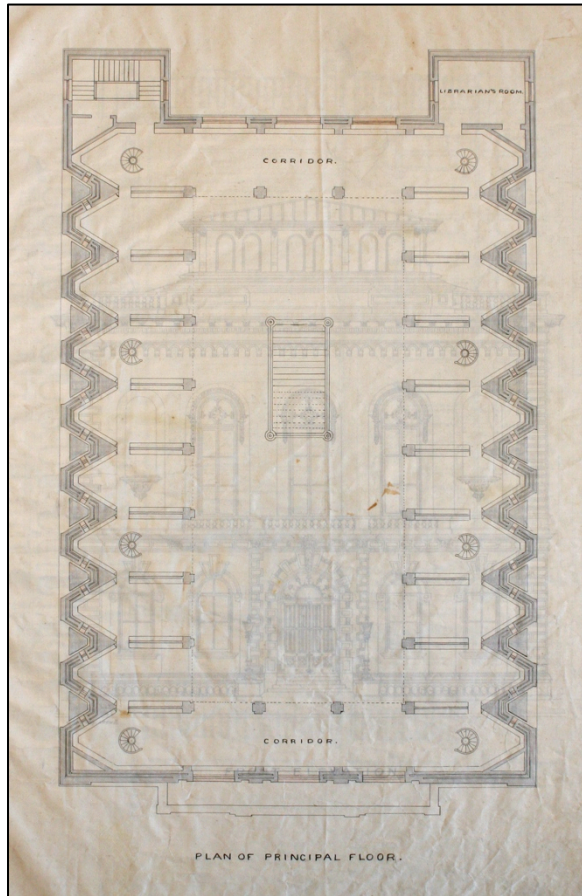


Figure 21: Plan of the Second Floor of the Boylston Street Building, Charles K. Kirby, 1857. Courtesy of Boston Public Library, Fine Arts Department. Photo by author.

This design gave many of the early philosophies of the library's leadership – motivations, intended users, visions for the collection – physical form. For example, the two levels of the library building directly reflected the leadership's two most favored user groups, as well as their prioritization of them: those in need (who were presumed to be largely casual readers, and not serious researchers, at least to start) were granted the easiest possible access, with all the popular materials and rooms in which to read them conveniently accessible on the first floor; serious researchers, in turn, were more welcome upstairs in Bates Hall, where the scholarly collections were kept, and which, though also public, was likely at least somewhat less noisy and crowded than the rooms below. Simultaneously, however, the placement of the research collections literally *above* the popular collections served as a visual and architectural metaphor for the process of

moral and intellectual uplift the library leadership hoped that the institution would generate among its users: as readers progressed through more of the popular collection below, they increased their chances of ascending in interest to the worthier collections

⁵⁴ Some of the physical details described here have been ascertained through examination of several photos of the building's interiors in the Boston Public Library's Flickr stream.

above (or so went the Trustees' thinking).⁵⁵

Despite the initial pride the library leadership showed in the building's practicality and functionality, however, it soon became clear that the structure was far from ideal. It would seem that rather than representing the pinnacle of utility that its champions had hoped for, it would look more like a failed prototype – informative for future design primarily through examination of its many flaws. The 1870 Examining Committee report, for example, cuttingly suggests that

It will generally be conceded, that whether we consider its external design or its internal plan, the building is equally a failure....[The] sins of the interior cry aloud continually and in vain for a remedy. The crypt-like Delivery Room, the narrow and ill-lighted Reading Room, the dark staircase with its wretched landings, where one stands groping for the handle of the door which should not be needed, the pretentious Hall above, fit enough for a music hall or an exchange, but as little like a library-room as it could well be made, the dark alcoves piled up three stories high, and shrouding the books in an almost impenetrable gloom; finally, the inexcusable absence of ventilation throughout the building; these are the daily and hourly misfortunes of all who have occasion to pass much of their time within these walls. To remedy the faults of the present building is impossible (as cited in Whitehill 1956, 79).

During his tenure as Superintendent, Justin Winsor would make valiant efforts to improve the utility of the space, going so far as to have additional windows punched through exterior walls. However, by the end of his time with the library in 1877, he too would write that the building's "faults are radical, and grew partly out of the inexperience of those, or rather a *majority* of the commission, superintending its erection...which induced an inability to comprehend the extent of work needful to be done in a rapidly growing Library, and partly from a sacrifice of fitness to a desire for ostentation" (as cited in Whitehill 1956, 79). Though one can imagine the by-then-deceased Everett wincing at Winsor's suggestion that aesthetics had trumped function in this case, Winsor's point regarding the inexperience of the building commission in library matters is suggestive. Though Winsor clearly intended it as a condemnation, I would propose that the inexperience of these early designers also reflected the extent to which they were acting as radical innovators in the library space. The commission on the building seems to have made a good-faith effort to

⁵⁵ One other, perhaps somewhat tangential, point of intrigue (and speculation on my part) is the ladies'-reading-room-cum-periodicals-room. The decision was made to repurpose the room based on the observation of decorum among the users in the Mason Street rooms; however, by the mid-1860s, decorum was beginning to break down in the new space, to the extent that the periodicals had to be moved behind the delivery desk, and disturbance of ladies using the reading rooms was being noted as an issue (e.g., "Report of the Examining Committee" 1865). One wonders, then, if there were any regrets about repurposing the room, or whether the absence of such a separate space influenced women's desire to use the library. The BPL's next building, in Copley Square, does not seem to have had a ladies' reading room either. However, there *was* a Children's Room in the new building, indicating a substantial shift in the thinking of the library leadership regarding the institution's user base by that time (Small 1895, 48-49).

make the building as useful for library purposes as possible; the problem was, given the lack of similar institutions from which to draw lessons, they had no idea what purposes a truly *public* library would ultimately serve, and thus what types of usage the building would need to afford. They thus included too little space for staff functions (e.g. binding and preparing books for shelving), too little lighting with which to retrieve books from the alcoves during the dark after-work hours, and too little ventilation to support the library's incredible public popularity. Though not all of these flaws were to be corrected in the next Boston Public Library building, the Boylston Street building provided a useful (if expensive) test case for the elements that future library architects would need to consider – and Justin Winsor clearly learned from this experience as he went on to become both Harvard's Librarian and the first president of the American Library Association (e.g., Whitehill 1956, 101, Brundin 1975).



Figure 22: Library Staff working in Bates Hall. Courtesy of the Boston Public Library, via Flickr.

These early errors in design, in turn, provide an interesting lens through which to view the contemporary cases included in this study. Arguably both Google Books and the Open Content Alliance suffered – and to some extent continue to suffer – from enormous initial structural flaws – the former tied to copyright, the latter to coordination and interface – however, the question becomes, will those flawed structures simply prove to be prototypes from which to learn and iterate, as in the case of the Boylston Street building? Or will they represent the final form of those projects – a sort of evolutionary dead end – while the world moves on to other, perhaps better ways of making books accessible online? These will be interesting questions to keep in mind in the chapters that follow.

4.1.2 Stacks

Within the physical structure of the Boylston Street building, there is one element that merits a somewhat closer look: the placement and control of access to the book collections. That is, all of the books, except for selected reference works, and in some periods, periodicals, were kept locked away from the public: behind the delivery desk on the first floor, and behind iron railings as well as up staff-only spiral staircases in Bates Hall. In later

parlance, this would come to be called a “closed stack” system.⁵⁶ For the BPL’s early leadership, however, this inaccessible “alcove” system was simply how library shelving was done. The concept of open shelving in libraries would not emerge until at least the 1880s, championed by reformers like John Cotton Dana (e.g., Dana 1897), and would not gain widespread acceptance until at least the early 20th century. Thus, the placement of the books out of reach of the public was not even a particular topic of conversation among the Trustees or others when discussing requirements for the building. The first mention of shelf access, in fact, appears in the Proposed Rules and Regulations for the Mason Street rooms, which specify that “no book shall be taken from its shelves at any time, except by an officer of the Library” – no exceptions (1853). After the library’s relocation to the more spacious Boylston Street building, Ticknor successfully lobbied for an alteration in this policy, to read: “No books shall be taken from the shelves in any part of the Library, by any person not employed in the service of the Library, *except such books as are deposited in the Reading Rooms for reference*” (*Proceedings of the Trustees 1852-1858*, September 13, 1858, emphasis added). This allowed library users to pluck items like encyclopedias and periodicals (in some years) from the Reading Room shelves at will – though still under the watchful eye of an attendant (*Proceedings of the Trustees 1852-1858*, September 9, 1858). In order to procure any other type of volume, however, patrons were required to submit written requests to the library attendants at the delivery desk (e.g., 1853).⁵⁷

Though the use of closed stacks in the Boylston Street building represented less of an explicit design decision than an instance of path dependency,⁵⁸ it nonetheless fit well with the leadership’s somewhat ambivalent assumptions regarding their user base. As noted earlier, there was a tension present in the motivations of uplift and character development held among the leadership: on the one hand, at least some of the Trustees, donors, etc., clearly had enough faith in the average Bostonian to warrant optimism about their ability to make use of the library for their own intellectual and moral benefit (e.g., “Report of the Trustees of the Public Library of the City of Boston” 1852, “Report of the Examining Committee” 1859). On the other, however, the very fact that they viewed their user base as *needing* intellectual and moral improvement suggests a certain level of paternalistic condescension (as most forcefully argued in Harris 1975a). The library was established to

⁵⁶ The term “stack,” in fact, never appears in any of the excerpts for this case, as it was not a term used to refer to library shelving at that time. The OED places the origin of that usage at 1879, in a *Library Journal* article by Charles A. Cutter (1.d.).

⁵⁷ As, indeed, is still the case for many materials at the BPL’s central Copley Square branch, including some of those retrieved for this research.

⁵⁸ That is, doing something in a particular way not because it is actually the best solution, but rather because it has already been developed as a solution in a different context (and possibly to a different set of problems), and because it would take considerable work to alter the surrounding processes in such a way as to make an alternative arrangement plausible. A more modern example of this phenomenon is the continued use of the QWERTY system for keyboards (David 1985).

provide “wholesome” and “useful” reading “for the intellectual, moral and religious progress” of the less fortunate (*Proceedings of the Trustees* 1852-1858, October 9, 1858), and its librarian was explicitly required to exercise discretion regarding the suitability of books for particular applicants (Board of Trustees 1853). In the documentation of the library’s planning process, there is a recurring undercurrent of low expectations regarding the conduct and abilities of library users, starting with Everett’s comment to Ticknor in 1851 that he would shy away from allowing lending because “Those who have been connected with the administration of such libraries are apt to get discouraged, by the loss and damage resulting from the loan of books” (as cited in Hillard, et al. 1876, 303), and flowing through to the 1865 Examining Committee’s comment that librarian assistance is necessary in part because “there are many persons so unskilful as not to be able to use advantageously a catalogue of the simplest construction” (“Report of the Examining Committee” 1865). The closed stack system physically instantiates these low expectations, barring the potentially questionable library user from laying hands upon valuable library collections without the explicit approval of a library staff member.⁵⁹

4.1.3 Catalogues

⁵⁹ It should be noted, for clarity’s sake, that these low expectations existed directly alongside the lofty optimism of library leaders with respect to their user base. As Wayne Bivens-Tatum has suggested, it should be recognized that “something as complex as the founding of a large public library could be motivated by multiple reasons, some of them perhaps contradictory” (Bivens-Tatum 2011, 111). Like most groups of humans, the founding BPL Trustees disagreed amongst themselves, and also sometimes seem to have individually held conflicting assumptions and goals. Though it makes for a less tidy story, it is nonetheless important to include elements of these conflicting sentiments in the analysis.

The final piece of library structure that bears analysis here is partially physical, but also more explicitly intellectual than the others: that is, the library's complex system of catalogues, developed first by C. C. Jewett and then further by Winsor to meet the needs of the BPL's unprecedented amalgamation of large and valuable collections with broad public access. The early BPL had roughly five different types of catalogue,⁶⁰ each serving particular purposes with regard to administration and/or access: (1) the Printed Alphabetical Catalogue, (2) the Alphabetical Card Catalogue, (3) the Accessions Catalogue, (4) the Shelf/Alcove Catalogue, and (5) the Slip Catalogue. The main catalogue, from a user interface perspective, was the Printed Alphabetical Catalogue, which started as a short title catalogue during the Mason Street years. As previously noted, in the years leading up to the opening of the Boylston Street building, the portion of the catalogue covering the Lower Hall collection was revised to a fuller form, with the Upper Hall catalogued separately later. This catalogue (or rather, each of these catalogues) was presented as a bound volume, copies of which were available in the library and also for sale for home use (Board of Trustees 1853). In contrast to the printed catalogue, the Alphabetical Card Catalogue was

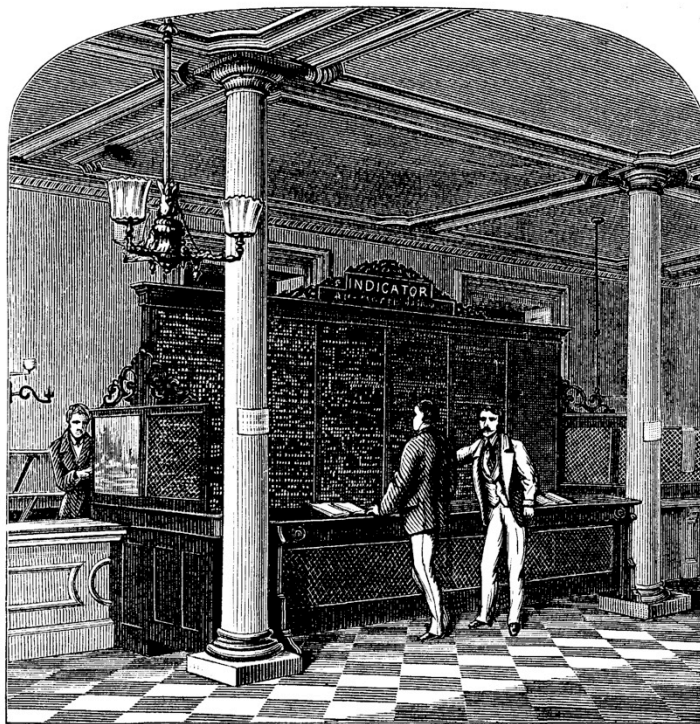


Figure 23: Delivery Room, showing the printed catalogues resting on an apparatus devised by Jewett to show whether books were checked out using colored wooden pegs (Winsor 1871, 153).

initially accessible only to library staff, and was ordered by title, as prescribed by the library's first proposed rules (Board of Trustees 1853). Before Jewett had completed the full version of the lower hall printed catalogue, consulting this card catalogue through a librarian appears to have been an important supplemental finding strategy ("Report of the Committee to Examine the Library" 1853, "Report of the Examining Committee" 1857). The Accessions Catalogue, in turn, provided a chronological record of purchases and donations, and served as the inventory of the library – and not so much as any sort of aid to navigation (e.g., "Report of the Examining Committee" 1856). The

⁶⁰ As the catalogues were called by different names in different periods, and as cataloging processes were constantly evolving at the BPL, the exact number is somewhat difficult to track.

Shelf/Alcove Catalogue⁶¹ was essentially a set of shelf-lists, which were physically adhered to the alcoves themselves, and which provided the names of the books in the order they were to stand on the shelves ("Reports of the Examining Committees" 1853-68). And finally, the Slip Catalogue was not truly a formal catalogue, *per se*, but rather a set of slips on which books were recorded in the order acquired. In that same order, the Slip Catalogue served as the rough draft for the Accessions Catalogue, while reordered by title, it became the rough draft for the Alphabetical Catalogues. It also served as the basis for an index to donations and benefactors ("Report of the Examining Committee" 1857).

These catalogues seem to have been developed to serve three major purposes: (1) to enable finding of books by different means, both by library users and by librarians; (2) to create an inventory of the books so that library staff (and through them, the Trustees, and ultimately the City Council) would know what the library had and whether anything was missing, and (3) to facilitate the retrieval of books from the alcoves by library staff. One could say that while the building provided the *physical* structure for storing and accessing the library's collections, the catalogues provided the *intellectual* structure. The library's need for this type of infrastructure, in turn, relates back to an element of the theoretical framework: the dual valence of communicative objects, such as books. Were the library a storehouse for lumber, or pencils, or coconuts, it would not require nearly as intricate a system for organizing and retrieving its contents. But books (and other documents), by their nature, have subject matter and provenance elements (e.g. authorship, time period, context of publication) that often have significant meaning for their interpretation; cataloging is one way of keeping those elements associated with physical materials, and, in the case of subject-based cataloging, of constellating the materials themselves into an epistemological framework that will help users to make sense of the collection. And with increasing scale, this scaffolding takes on yet more significance – something the Trustees themselves noted in their preface to the printed catalogue, in 1858:

Next to the collection of its books the Trustees look upon this catalogue as the most important part of the Library, for it is the part by which the whole mass of its resources is opened for easy use; -- the key by which all its treasures are unlocked to the many who, in this community, are now asking for them so often and so earnestly. A large Library without good catalogues has sometimes been compared to a Polyphemus without an eye, and more frequently to a chaos, which it certainly too much resembles. This reproach the Trustees hope to avoid for the Public Library, which they desire above everything else to render useful (*Proceedings of the Trustees 1852-1858*, October 9, 1858).

As the Trustees note, the large size of the library was a driving force behind the need for all these different catalogues. A small set of volumes, after all, can be organized in any number

⁶¹ These seem to have been called the Shelf Catalogue until 1855, and the Alcove Catalogue thereafter ("Reports of the Examining Committees" 1853-68)

of ways – by author, by title, by color – or even not organized at all, with little effect on the collection’s usability. However, once a collection reaches the point of filling several shelves, and especially once it fills several ranges of shelves, spread out over multiple rooms, an extensible, scalable system of organization becomes of fundamental importance. Else, as the trustees note, chaos ensues. At the scale of the BPL in the 1850s and into the 1860s, the system described above sufficed to keep the chaos more or less in check. However, as the collection grew, it became clear that bound catalogue volumes would no longer serve: the difficulty of adding to them made them quickly go obsolete, and their linear structure provided limited options for alternate orderings of the works they contained. For these reasons, in 1871, Winsor began work on the BPL’s first public card catalogue (Whitehill 1956, 97) – a highly extensible, scalable system that remains in use in certain BPL departments to this day, in parallel with online versions developed later (though generally containing materials not catalogued in the online public catalogue).

4.2 Processes

The previous section described material, physical forms of constraint on library use (where constraint is understood as both restrictive and enabling); however, less embodied forms of constraint were in operation as well. The final section of this chapter will thus explore a different set of structures at play in the library: that is, the *social* structures, which I will call “processes,” established and perpetuated through the purposeful actions of early BPL administrators in agitating for political changes, writing library rules and policies, and generally carrying on the work of the institution. The subsections here will look quite familiar to anyone acquainted with the departmental structure of any given modern library – Management, Acquisitions, Cataloging, Public Service, and Protecting the Collections⁶² – and some comparisons will be offered. Then, as now, these represented the basic clusterings of library functions, and thus they also suggest convenient thematic divisions of the processes and policies enacted during the early administration of the BPL.

4.2.1 Management

The management of the library, especially at the top levels, was a frequent point of contention between the Trustees and the City Council, each of whom claimed decision-making priority. Additionally, as it evolved, the library’s administrative structure increasingly resembled that of the higher education institution it aspired to be: an enormous collection of committees overseen by an intellectually accomplished director selected and retained according to policies explicitly modeled upon academic tenure.

⁶² I say “Protecting the Collections” instead of “Preservation” because in modern library contexts, “preservation” generally encompasses physically repairing, modifying, or digitizing books, in addition to shielding them from rough use. In this chapter, however, I will be focusing specifically on the task of protecting books from their (potential) users, via access restrictions and punitive measures.

4.2.1.1 *Leadership Politics*

The office of Librarian was created along with the institution, by city ordinance – and that same ordinance further specified that the selection of said librarian, as well as the length of his tenure and amount of his compensation, would be under the control of the City Council (Boston Common Council 1852a, 3). It was thus the City Council that appointed Edward Capen as Librarian⁶³ – a position in which he would remain for more than two decades, until 1874 (Whitehill 1956, 26, 54). Still, based on the original ordinance, the Librarian had to be appointed (or reappointed) annually by concurrent vote of the City Council, which meant that the position was subject to becoming a political chess piece. As the library's collection and usage grew, and especially as the institution prepared to move into its first dedicated building, the Trustees took increasing exception to the tenuousness of its leadership. Additionally, by that time, as noted previously, C. C. Jewett had already been demonstrating his excellence as a leadership candidate through his remarkable work on cataloging the second Bates donation. Thus, in 1857, the Trustees petitioned the City Council for an alteration in management structure, creating the position of Superintendent to take over most of the higher-order functions previously served by the Librarian, and effectively demoting the Librarian to what we would now call the head of circulation (Board of Aldermen 1857, citing Everett). In addition, they also argued that the selection of this individual should rest in their hands, rather than those of the City Council; in their reasoning for this alteration, their conception of the library as an educational institution is quite explicit:

The Trustees conceive that [appointment by the City Council] is too precarious a tenure for such an office. The place of Librarian in a great public library nearly resembles that of a professor in a seminary of learning. The Trustees are not aware that it has ever been deemed expedient, in any part of the country, to subject the teachers or the librarians in our universities and colleges to the uncertainty of an annual election, by public bodies partaking largely of a political character. As the Trustees are directly responsible to the City for the condition and working of the institution, and as the duty of making the requisite regulations for its management, and of seeing that they are carried into effect, devolves on them, they are of opinion, for obvious reasons, that the appointment of the Librarian and of any other officer, who may be established in pursuance of the foregoing recommendation, should be devolved upon the Board (Board of Aldermen 1857, citing Everett).

While the Trustees would succeed at that time in creating the position of Superintendent and in placing Jewett in the role, however, they would not be granted full control over the

⁶³ Though interestingly, another influential leader in the public library movement, William F. Poole, very much wanted the position, and would repeatedly vie for it before moving on to serve as the founding librarian of both the Cincinnati and Chicago Public Libraries, in turn. Interestingly, Whitehill suggests that Poole's chances at the job were hampered by his having attended Yale, rather than nearby Harvard (Whitehill 1956, 26, 76).

selection process until nearly a decade later. In the Examining Committee report for 1865, this shift is applauded using much the same educational parallel provided earlier, suggesting that the “office of Superintendent of a large, and constantly increasing, public Library is one of the highest and most honorable in the community. It is on a par with that of the Presidency of a University.” The Examining Committee, however, goes further, explaining that

more is required of a superintendent than can be found in most men. He must be a man extremely learned in books; a proficient in the modern languages; a competent and skilful business agent, and one of infinite tact, to enable him to meet, with equanimity, the various annoyances incident to the position. In addition to these natural and acquired qualifications, it would be unfortunate if he should not have studied in the various large libraries of Europe, in order that he might avoid their dangers, or seek to engraft their good qualities upon the institution committed chiefly to his care. It is obvious that such a man would not often be found in or near the arena of mere politics. Nor, having chosen his profession of librarian, would he ever for a long time submit to the annual uncertainties of political strife in the matter of his election (“Report of the Examining Committee” 1865).

To the early Trustees, the Boston Public Library was meant to be the university of the people; it thus makes sense that they should view the Superintendent as its chief academic officer.

4.2.1.2 Committees

The relevance of committees to the administrative structure of the BPL can be stated quite briefly: then, as now, there were many. I would suggest that this is indicative of the library’s pedigree as both a public institution and an educational one, both of which incline toward committee rule even to this day. Within the Boston City Council, committees were established initially to found the library as a city institution, and later to plan for the building (Boston Common Council 1852b, Boston Common Council 1854b, Boston Common Council 1854a). Additionally, the library’s Board of Trustees was effectively a committee itself, on which the Mayor and two City Council members sat by city ordinance (an ordinance which also required the annual formation of an Examining Committee) (Boston Common Council 1852a). And finally, within the Board of Trustees, committee formation ran wild, not only tasking pairs or triads of Trustees to execute or manage particular library functions – e.g. administration, the building, “purchase and binding of books” – but also sometimes creating committees of only one member, for example to confer with a potential donor about his or her donation (*Proceedings of the Trustees* 1852-1858, February 12, 1858).⁶⁴ Notably, though this structure seemed fairly functional as an

⁶⁴ And, of course, there was also a Committee on Committees, to help sort all this out (*Proceedings of the Trustees* 1852-1858, February 12, 1858).

administrative strategy for many years, it seemed to work better as a background to strong leadership than as a primary strategy in itself: when there was a Ticknor, an Everett, a Jewett, a Winsor, the direction of the institution as a whole could – and did – follow their lead, whereas in periods like the one following Winsor’s departure, the institution seemed to persist more on ever-diminishing momentum than according to any systematic strategy (Whitehill 1956, 112-17). Indeed, one of Winsor’s former lieutenants, Frederic B. Perkins, would write to the Mayor during this period that the library was “not governed, but fumbled,--by a set of committees” (as cited in Whitehill 1956, 116-17).

4.2.2 Acquisitions

Much of what is interesting about the BPL’s acquisitions process has already been covered in other sections: the mix of donations and purchases; the personal, high-level attention required to increase the library’s scale (e.g. Ticknor’s book-purchasing trip to Europe); and the prioritization of collecting and organizing popular materials over scholarly ones in the early years. Going much further than this in examining the nitty-gritty specifics of acquiring books in the mid- to late-nineteenth century will not be very useful. However, there is one point of interest in the acquisitions procedures that still merits further note: that is, the input sought out by those responsible for acquiring the library’s early collections, from both local domain experts and the users themselves.

Prior to Ticknor’s European purchasing trip, he attempted to avoid having to make the voyage by working with Jewett to assemble comprehensive bibliographies in various subjects and sending them along to Bates in London. In order to do this, he started by requesting lists of books in particular subject areas from “literary and scientific gentlemen of known eminence” (“Fifth Annual Report of the Trustees of the City Library” 1857, Superintendent’s Report). As Ticknor’s first biographer, George Hillard, would note, these requests garnered

contributions of much consequence from such men as Professors Agassiz, Bond, Cooke, Felton, Hayward, Holmes, Lovering, Pierce, and Dr. John Ware; from Professor W.B. Rogers and Judge Curtis; from Colonel Thayer of the Army and Captain Goldsborough of the Navy; from engineers and architects, clergymen and men of letters.

And, Hillard goes on to explain,

With these, and with all the bibliographical resources they could command, Mr. Ticknor and Mr. Jewett worked, in Mr. Ticknor’s library, for more than two months, Mr. Jewett remaining there eight hours a day, preparing the lists that were to be sent to Mr. Bates. These lists, embracing above forty thousand volumes, were successively forwarded, and were approved by Mr. Bates (Hillard, et al. 1876, 310).

Though Ticknor did not manage to avoid the trip, these lists – and others he later requested be made while still in Europe – nonetheless provided the outline for the library’s purchasing goals with respect to the second Bates donation, guiding both Ticknor himself and the various purchasing agents with whom he and Bates contracted throughout Europe (Hillard, et al. 1876, 311-12, Whitehill 1956, 49).

Perhaps more intriguing, however, was the stance taken by the Boston Public Library on user input with regard to the collections. That is, the library went out of its way to offer patrons the opportunity to request the purchase of books that they wished to read which were not already held by the library. This, again, shows Ticknor’s early influence – and as Hillard explains, the policy was neither uncontroversial nor immediately effective:

a favorite proposal of [Ticknor’s] was much discussed and somewhat opposed among the Trustees, -- that of allowing frequenters of the Library to ask for books to be purchased, and for that purpose to supply cards or blanks for such applications. He gained this point...and persevered in having it not only offered but urged, although for ten years this great and useful privilege was not appreciated. Until 1865 the public could not be induced to understand or avail itself of this opportunity, and, before that time, the Trustees had come fully to apprehend the value to them of such requests, in pointing out what was desirable to purchase, and would be immediately useful (Hillard, et al. 1876, 307).

Indeed, by 1866, the Examining Committee would begin to evince some despair of ever gaining the public’s involvement in collection development, going so far as to declare that “if books, such as are wanted, are not to be found in any of its various divisions, it is, in no small degree, the fault of the public,” due to their failure to avail themselves of the opportunity to request purchases (“Report of the Examining Committee” 1866). Still, the Committee does go on to note, with a thin shade of hopefulness, that “the last year ending August 1, 1866, by uncommon exertions and personal application [the number of requests] was raised to three hundred and six [from less than sixty in prior years], a number still greatly less than it should be, but which it seems very difficult to increase” (“Report of the Examining Committee” 1866). Despite these difficulties, the idea of seeking this kind of input had clearly obtained strong support among the library’s leadership by this point: in that same 1866 report, there is a renewed call for library users to request whatever they might wish to read, as an important aspect of what makes the BPL a *public* institution, rather than a closed-off collection like that of the British Museum or the Vatican. The Boston Public Library, after all, was established not to gather up and protect rarities, like those institutions, but to serve a set of purposes that all require both public access and public *engagement*: education, political and moral uplift, innovations in technology and business – all of these occur only with the consent and involvement of the library user. The policy of

attempting to engage users in collection development meshes well with the promotion of engagement required to meet these other goals.

4.2.3 Cataloging

As with acquisitions, much of what is interesting about the cataloging processes in the early BPL has already been said in a previous section, in this case, on Structures. However, it is worth noting that while that section described *the catalogues* as an intellectual support structure for organizing and gaining access to the collections, the actual process of *cataloguing* implicates a slightly different set of issues. In particular, anticipating some of the later cases covered by this dissertation, the BPL's experiences clearly illustrate the way in which metadata creation becomes simultaneously vastly more important and vastly more difficult to do where an exceptionally large-scale collection is concerned. It also illustrates, to a certain extent, the influence of available technology: all cataloging in the early Boston Public Library – into all of those many catalogues – had to worked out and at least drafted by hand, which made it a relatively slow and intellectually rigorous process. In his 1857 annual report, Jewett provides a detailed account of the process utilized during the most difficult years for the catalogers – those in which the Bates donation was flowing at a mad gush into the library's possession, despite the institution's general lack of space to accommodate them at the time. Figure 24 illustrates the flow of this process in simplified terms. In concluding this description, Jewett notes:

The processes are indeed numerous, through which every book must pass before it can be incorporated among the useful possessions of the public library. Some of these processes consume much time. It has, for instance, been found by long experience that, on an average, not more than thirty or forty titles a day, or about 10,000 year, can be accurately written by the most competent and diligent person. But it is gratifying to note that when this work is once done, it is done for all time, so far as these books are concerned. ("Fifth Annual Report of the Trustees of the City Library" 1857, Superintendent's Report, 36).

Jewett's suggestion that the cataloging had been accomplished "for all time" may have been somewhat optimistic; particularly as library cataloging systems evolved over the next fifty years – and as the BPL moved into a new building – alterations to the system and some degree of re-

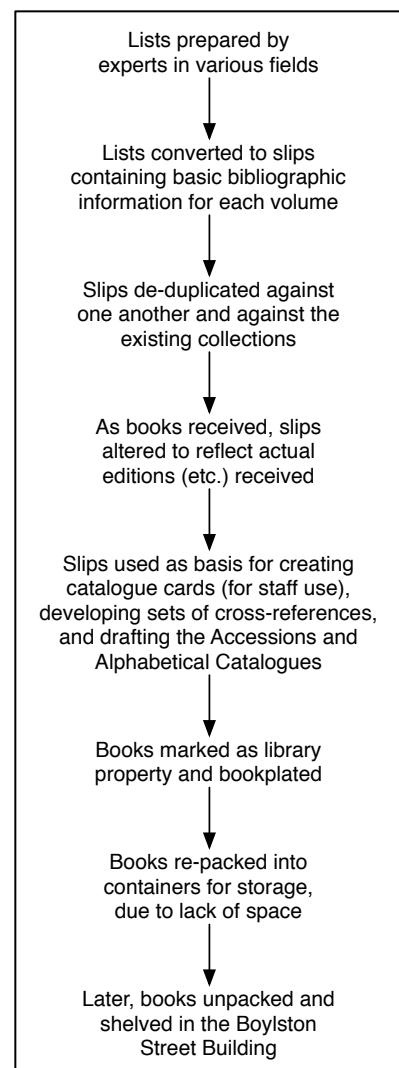


Figure 24: Cataloging process for second Bates donation, 1856-57 ("Fifth Annual Report of the Trustees of the City Library" 1857, Superintendent's Report 32-38)

cataloging would become necessary. However, the challenges involved in the BPL's cataloging process in the late 1850s were substantial, in no small part because of the explosion in the collection's size at that time – indeed, the complexity of the task formed one of the key justifications for the creation of the Superintendent role in the mold of an academic official (Board of Aldermen 1857, 6-7). As time went by, the process grew increasingly systematized: a shift that I will suggest is a key requirement of doing most anything of this sort on a large scale, especially in a way that can be sustained over time. Indeed, this point will recur in every case discussed in this dissertation. Without a certain level of standardization and routinization of administrative processes, not only does the construction of a large-scale information resource proceed more slowly, but the loss of a single individual can wreak havoc on the entire arrangement. Still, in the early years of the BPL, those systems were still being negotiated – and looking back from today's perspective of “big data” initiatives and digital libraries, there is perhaps something comforting in the recognition that the creation of metadata on a large scale was at least as difficult for the scholar-librarians of the early BPL as it is for the computer-assisted systems of the present.

4.2.4 Public Service

Like most elements of administration in the early Boston Public Library, the definition of “public service” evolved over time. On the most basic level, a high priority was given to keeping the building open and books available for retrieval for a lengthy period of every “secular” day – ideally, according to the 1852 Trustees' Report, “as many hours of the day as possible, and always in the evening” (“Report of the Trustees of the Public Library of the City of Boston” 1852, 19). Accordingly, even in the library's modest rooms on Mason Street, the Reading Room was opened to the public from 9 am to 9:30 pm, with book delivery available between 3 and 8 pm (1853). After occupying the Boylston Street building, the open hours were extended later, to 10 pm, while the book delivery hours were extended earlier, beginning at 10 am each day. In the lower hall, book delivery was still available until 8 pm; however, in the upper hall, where the lighting was exceedingly poor, book delivery was only offered “until gas-lighting” (*Proceedings of the Trustees* 1852-1858, September 13, 1858). The extension of the library's service hours this far into the evening, as well as to Saturdays,⁶⁵ was a very explicit reflection of the Trustees' desire to attract working people to the institution, as it allowed them to stop in after the work day had ended. By way of comparison, it is worth noting that New York's Astor Library was during the same period only open from 10 am until 5 pm – and its collections did not circulate (Lydenberg 1916, 568).

Beyond the hours of physical access to the library and its collections, however, library users were also offered the assistance of staff for purposes which developed and expanded over

⁶⁵ Sunday opening would be introduced – amid a huge amount of controversy – in 1889 (Wadlin 1911, 103).

time. In the very earliest years, the main point of interaction between staff and user was at the level of circulation services: retrieving requested books from the closed stacks and allowing them to either be used within the building or taken from it for home use. However, as the institution began to mature, the primary documents reveal increasing discussion of staff functions that look more and more like modern reference services – despite the fact that the first articulation of such phenomena in libraries is generally dated decades later, credited to Samuel S. Green’s article “Personal Relations Between Librarians and Readers” (Green 1876, Janes 2003). This section will discuss each of these types of service in turn.

4.2.4.1 *Circulation*

From the perspective of modern libraries, the actual lending of books can seem, well, somewhat boring; at the very least, it is certainly taken for granted. However, as noted previously, this was not yet the case in mid-nineteenth century Boston. Even between Ticknor and Everett, the question of whether the average citizen could be trusted to care for publicly owned library volumes outside of a supervised setting was a debate rather than a given (Hillard, et al. 1876, 303). Had Ticknor not prevailed in his conviction that the library collection must circulate, one wonders what not only Boston’s public library, but American public libraries in general, might have looked like, given Boston’s influence. As it happened, Ticknor did prevail, and circulation became perhaps *the* major business of BPL staff for decades to come. Indeed, up to 1859, every mention of service to users in the primary documentation refers to circulation services – fetching things from the closed stacks and charging them out to borrowers (e.g., 1853, "Fourth Annual Report of the Trustees of the City Library" 1856, Board of Aldermen 1857, citing Everett, "Seventh Annual Report of the Trustees of the City Library" 1859). The dominance of this type of work in the daily tasks of the library staff is vividly illustrated in the Trustees’ 1857 request for a Superintendent:

The work now done in the library, in addition to that which is done by the Trustees themselves and their Committees, is performed by the Librarian, and by one male and three female assistants. The most laborious part of this work consists in the daily attendance for five hours upon those who resort to the library to borrow and return books; to the number sometimes of five or six hundred, in one afternoon. It will readily be conceived that the greatest promptness on the part of the entire force employed in the library is required in order to wait upon so many persons in a satisfactory manner. When the number of books is greatly increased, the demand upon the circulating department will no doubt be increased in the same proportion; and there is reason to suppose that, not merely five hours daily, but the whole time of the Librarian and his present number of assistants, may be needed for the delivery of books, and the

reception of those returned (Board of Aldermen 1857, citing Everett).⁶⁶

In part, the laboriousness of the circulation function of the library derived from the closed stack system, regardless of usage level: if users cannot simply walk up to the shelf and pick up their own book, staff time will be taken up doing that for them. To a greater extent, however, the immense scale of this task was dictated by the somewhat unexpected and immediate popularity of the institution. “Five or six hundred” book requests in a single day is a huge number, particularly when retrieving any given volume might require ascending one or two stories on a spiral staircase, circumnavigating Bates Hall, and/or climbing up onto a ladder in a dimly lit book alcove. Such a system might be practical for a very popular library with a small collection, or for a lightly-used library with a large collection, but for an exceedingly popular library with an enormous collection, it would soon become unwieldy.

4.2.4.2 “Reference”

The idea of library staff as intellectual guides to the collections did not fully take hold until a bit later, as noted. However, by 1857, the BPL Trustees would begin to make moves in this direction. In the same request to the Aldermen cited above – and thus directly alongside the contention that circulation takes a huge amount of staff time – the Trustees suggest that more in-depth service needed to be provided to library users, and that this need indicated the importance of creating a more faculty-like role at the head of the institution:

An important part of the duty of those charged with the management of large public libraries is to attend to persons, both citizens and others, who resort to them for the purpose of scientific and literary research. Many persons will visit the Public Library in Boston, for this purpose. It will contain very many valuable works of reference, and books too costly and rare to be put into circulation, but which will be consulted by those who visit the institution. It belongs to the management of a great public library to answer the inquiries and to facilitate the researches of persons of this class, and no small part of the time of some of its officers will be taken up in this way. An extensive knowledge of books, of ancient and foreign languages, and of science and literature generally, is indispensable for the performance of this duty in a satisfactory manner (Board of Aldermen 1857, citing Everett).

The Superintendent, in the Trustees’ mind, was to be a scholar-librarian, broadly enough educated to be able to claim competence in discussion and instruction regarding large portions of the collection, and thus able to help even serious and scholarly patrons to navigate them – a mold which both Jewett and Winsor fit to a tee. And indeed, as the library turned increasingly toward developing its research collections in the late 1850s and

⁶⁶ It is worth noting that this particular passage had the dual purpose of highlighting the labor-intensiveness of circulation and providing a graceful excuse for effectively demoting the current librarian, Edward Capen, who would retain responsibility for only the parts of the library to do with circulation as the newly-created Superintendent role would assume the actual high-level management of the institution.

early 1860s – and also, one suspects, as the library leadership gained more understanding of the actual capacities of their user base – so too increased the emphasis on the need for this type of guidance as a necessary complement to the catalogue. By 1865, the Examining Committee would note:

The Library is now large enough to be made in every manner useful to the student. Two things are important to this end, viz. : First, catalogues, well-arranged, complete, and accessible; and, second, an attendant, by whose assistance the resources of the Library, not attainable from the catalogues, may be known. ...[There] are many persons so unskilful as not to be able to use advantageously a catalogue of the simplest construction, and whose inquiries, therefore, would not be met by anything of the kind. There are others, on the contrary, quite able to use with skill any means in their power, and yet, when entering upon any special subject of investigation, they may be at a loss what books to ask for. Now, to both of these classes an attendant might be able to make valuable suggestions. This attendant should be one of the higher officers of the institution, a courteous and able man, who, from his well-stored mind and thorough acquaintance with at least the outlines of the varied learning contained in the Library, might be able to suggest to a student some of the books most likely to aid him in his investigations ("Report of the Examining Committee" 1865).

The Committee goes on to add that this responsibility would naturally devolve upon the Superintendent – and often did, at that time – but that given all of that officer's other responsibilities, it may be advisable to create a new position to serve this purpose. Although no action seems to have been taken on this recommendation at that time, the suggestion alone is historically notable, as it essentially represents a proposal to create the first dedicated reference librarian position, long before that term would come into usage in the context of personnel,⁶⁷ and also more than a decade before Green's similar suggestions on the pages of *Library Journal*. And despite the apparent absence of any dedicated staff member for this purpose over the following decade, Superintendent Winsor made great strides in the area of reader's advisory, for example by appending notes suggesting starting points and potential directions for inquiry directly into the catalogue, under the subject-references for history, biography, and travel for the Lower Hall collection. In doing so, he would later write, his goal "was not to propound positively any course of reading, for there is danger always in dogmatism, however right its foundation may be," but rather "to render the ordinary reader more able to choose to his liking when an undistinguishable mass of equivalent titles perplexed him" (Whitehill 1956, 84, citing Winsor, 1877). As an element of its mission to serve both casual readers and serious researchers, the Boston

⁶⁷ The use of the modifier "reference" to describe *books* – generally ones that do not circulate and are kept on open shelves within the library reading room – goes back much further, as should be evident in several passages from primary sources here.

Public Library thus seems to have positioned itself as an innovator in the guidance and assistance of library patrons well before it became fashionable to do so.

4.2.5 Protecting the Collections

In Bates's initial letter offering a donation to the library, he specifies as a condition that the institution "shall be perfectly free to all, with no other restrictions than may be necessary for the preservation of the books" (Trustees of the Public Library 1865, letter to Mayor Seaver, October 18, 1852). This requirement – and its accompanying caveat – would guide the development of a variety of structures and practices in the early BPL. The ideal of broad access was ever-present, in the building's open hours, in the library's expansive collection priorities, and in the explicit welcoming of individuals from a wide range of classes and backgrounds as users. However, the caveat to that ideal – that the books must be protected – was equally pervasive. Some of the ways in which this caution emerged have already been discussed: the storage of the books in closed stacks, available only at the discretion of library staff; works of particular rarity or value, as well as several donated collections, not being allowed to leave the library building; the limitation of the scope of the user base mainly to Boston taxpayers. However, there are two additional sets of policies put in place to protect the collections which merit mention: first, the restriction of access to library use and borrowing based on age, reputation, and conduct, and second, the measures taken to punish abuse of materials.

4.2.5.1 Access Restrictions

The restrictions placed on who was allowed to use the collections of the library are most clearly enunciated in the Rules and Regulations of the library, as first proposed in 1853 and revised over time after that. As noted above in the discussion of users, these rules granted borrowing privileges to several specific classes of individuals, such as clergy and city government officials, and were quite liberal for the time period. However, three restraints placed on access by these rules – based on age, demonstrable reputability, and conduct – are noteworthy both for their thematic pervasiveness even beyond the rules documents and for their relationship to the early leadership's nervousness about whether the entire public could actually be trusted to use library materials without causing them undue harm.

4.2.5.1.1 Age

One very basic restriction on library use was the minimum age requirement. As of 1858, in order to use the reading room, patrons had to be over age 16; in order to borrow, over age 18 (*Proceedings of the Trustees* 1852-1858, September 13, 1858).⁶⁸ Children were allowed in the library, and were even allowed to retrieve books for adult cardholders; however, they were

⁶⁸ The reading room age restriction appears in the 1853 rules also, but there is no apparent mention of a separate age restriction for borrowing before 1858.

generally seen as more of a nuisance than a welcome set of users. Indeed, the 1865 Examining Committee suggested that children retrieving books risked damage to the collections, and even went so far as to recommend that “Small children, except in company with parents, should be totally excluded, and, in general, every noise that can be avoided should be prevented” (“Report of the Examining Committee” 1865). The sole exception to this general distaste for the presence of minors in the library was the policy of granting library privileges to especially high-performing students: winners of the public schools’ Franklin Medal in particular, but also set numbers of students recommended by each of several local schools in each year (*Proceedings of the Trustees* 1852-1858, September 13, 1858, 1853, Board of Trustees 1858).

From a modern perspective, the idea of a public library where children were generally unwelcome seems almost bizarre. After all, the children’s sections at current public libraries are often the most beloved and heavily utilized portions of the institution. However, for the early BPL, this policy was directly in keeping with two guiding institutional principles. First, the central purpose of the institution, as envisioned by its founders, was to continue education beyond the school years; as such, it was envisioned as a place for adult reading. Children, in this view, were already being served, because they already had the benefit of the public schools; it was young adults who lacked resources for continuing their education. And second, excluding children was seen as a key way of protecting the collections. After all, if the leadership was uncertain about whether local *adults* could be trusted with the care of the library’s books, how were they to rationalize the extension of access to seven-year-olds? By the time the Copley Square building was erected, this thinking had shifted enough that the new building included a Children’s Room (Small 1895, 48-49); however, in the period under consideration here, children remained very explicitly unwelcome.

4.2.5.1.2 *Proof of Personal Worthiness*

Even among adults, specific criteria existed for establishing sufficient trustworthiness to make full use of the library – in particular, either by presenting acceptable proof of personal character, or by offering a monetary deposit on the items borrowed. In order for borrowing privileges to be granted *for free*, individuals had to be personally vouched for in some way. This could mean being personally known to the librarian, being recommended by a “responsible citizen,” holding a position in the city government or clergy, having made a donation to the library of over \$100, or being one of the above-mentioned types of honor students. If none of these conditions were met, however, “inhabitants of the city” could often still borrow books – just not for free. In that case, the individual would have to provide a deposit equal to the full value of the book, or equal to the value of the whole set, if the volume is part of a set (1853, Board of Trustees 1858). One imagines this could have been a fairly prohibitive requirement for many potential users; however, it is unclear how

many people it actually excluded, as no statistics appear to have been kept on the need for or utilization of this option.

This set of restrictions on access is intriguing. It seems possible that the requirement for personal references (or particular kinds of social status) may well have excluded some of the very users the library was explicitly intended to serve – particularly individuals without the resources to purchase their own reading material. These individuals, by definition, would not have been able to place a full-value deposit on a book. Whether they would have been able to secure the necessary character reference for free use of the library is less clear. As noted, there do not seem to have been any statistics on this topic collected during the time period in question, and thus any speculation about the height of this hurdle is just that and no more. But it is nonetheless intriguing to ponder – who might have been able to serve as a reference for the average working person? Or the average immigrant? And would they have agreed to do so? It is possible to imagine types of individuals who might well have vouched for working-class people – religious leaders, work supervisors, etc. – so perhaps this was not a significant problem in actual practice. Still, in the absence of further information, the possibility that this *was* an obstacle is worth raising as a potential countervailing force to the library's goal of providing access to those in need.

4.2.5.1.3 *Conduct*

Once the requirements of age and respectability had been met, actual user conduct formed the next most fundamental constraint on access provision. Indeed, one of the very first criteria listed for access to the library in every version of the rules produced in this period was that users should be “of such orderly conduct and condition as not to interfere with the occupations and comfort of others” (1853, 3, Board of Trustees 1858, 3). Initially, the library's leadership showed great faith in their patrons' capacity for such good behavior, for example noting that:

Great reliance has thus far been placed, and not without reason, on the good feeling and moral sense of borrowers, as the surest protection from injuries of both kinds [that is, loss of and damage to books]. Should this reliance fail, and should it become necessary to resort to stringent measures of detection and repression, the Trustees will confidently look to the support of an intelligent and virtuous community ("Fifth Annual Report of the Trustees of the City Library" 1857).

And indeed, as of 1858, this faith seemed to have been warranted, with Jewett reporting that he was

happy to be able to state that since the building has been open to the public, the most perfect decorum has been observed by all its numerous visitors. Not an expression of dissatisfaction or impatience has been reported to me. No better evidence could be given of the high value the public place upon the great advantages which are here so freely offered, and of the safety of opening such an

establishment without other restrictions, than such as are required to distribute its benefits equally among the greatest number ("Sixth Annual Report of the Trustees of the City Library" 1858, Superintendent's report).

Still, this observation of "perfect decorum" would not hold true through the Civil War years. By 1866, the Examining Committee would note that theft and destruction of materials had become a troubling habit among library users, suggesting that the poor state of the Lower Hall collection was

in consequence of a great abuse of its privileges by those to whom it was opened; in consequence of the shameful mutilation and defacement of its books; and in consequence of their disappearance in such numbers that it seemed as if those who borrowed them were, from the imperfection of the record and of the administration, invited to wrong-doing.

And furthermore, the same report also notes damage and loss in the periodicals collection:

in 1864 and 1865, Reviews and Magazines in considerable numbers were stolen, as a few had been earlier; many were mutilated; and still more were abused in various ways, sometimes by gross words written on them, and sometimes by indecent drawings, offensive not only to the many young ladies and matrons daily frequenting the room, but to any person of becoming manners and education ("Report of the Examining Committee" 1866).

Partially as a result of this shift in user behavior, a new system was introduced for tracking lending, which would more closely tie instances of abuse to those responsible, and the periodicals were removed from open shelving in the Reading Room to storage behind the delivery desk. These issues indicate one level on which the tension in the Trustees' thinking regarding users – the desire to provide access paired with mistrust for the common man – would be negotiated in practice, as the behavior of at least some users began to validate that mistrust.

And indeed, this tension continues to be negotiated in public libraries. Of the restrictions described here, the one based on conduct seems closest to what would be considered typical now: if a patron is disruptive to other users, she will still generally be ejected, and if the behavior continues, she will still have her privileges rescinded (e.g., Ann Arbor District Library 2008, Seattle Public Library 2009). Still, it is worth noting that while today's behavior rules for public libraries can seem a little frighteningly geared toward preventing actual violence and use of weapons and drugs, in the mid-nineteenth-century BPL, the goals of the conduct restrictions were more modest, geared simply toward maintaining a decorous atmosphere and preventing vandalism and theft.

4.2.5.2 *Punishing Abuse*

The final piece of protecting the library's collections is really just the flip-side to the conduct restrictions just discussed. That is, when a patron defied the rules of the institution,

punitive action was taken. The procedure for reporting such delinquency is outlined in the first proposed rules of the institution:

Any person abusing the privileges of the Reading Room by unbecoming conduct there, or by the violation of any of its regulations, shall be at once reported by the Librarian to the Trustees, and by them excluded from its benefits for a time, or permanently, according to the nature and degree of the delinquency or default; but in case of any gross offence the Librarian shall act summarily in the matter, and cause the offender to be at once excluded from the room, reporting the case to the Trustees as soon as possible, in writing, for their final decision (1853, 4-5).

And although that passage applies specifically to misbehavior in the Reading Room, similar penalties awaited those who abused the privilege of borrowing books (1853, 9).

Additionally, then as now, borrowers who were late in returning books were subject to fines: three cents per day for especially valuable items, two cents for most others (*Proceedings of the Trustees* 1852-1858, September 4, 1858).

Still, sometimes the Trustees would go further in penalizing delinquent borrowers – considerably further, in fact, than most libraries would go today. For example, in 1856, the library published the names of six delinquent borrowers in the local newspapers (*Proceedings of the Trustees* 1852-1858, March 18, 1856). And any book kept more than a week past its due date was subject to being “sent for at the expense of the delinquent” (*Proceedings of the Trustees* 1852-1858, September 4, 1858) – that is, the library would send someone to the borrower’s home to retrieve it. Finally, the 1865 Examining Committee report suggested that given the increase in abuse, perhaps the library should involve the local police in tracking down delinquent borrowers and holding them accountable (“Report of the Examining Committee” 1865).⁶⁹ These forms of recourse for delinquency make somewhat more sense when one considers two factors. First, at the time, Boston was still relatively small in population. The same interconnectedness of personal relationships that made the requirement of personal references a practicable policy would have also facilitated the tracking down of delinquents, who were likely often personally known to the library’s administrators. And second, though books had generally become considerably cheaper with advances in printing technology in the nineteenth century, those collected by the library were often still relatively costly to replace. Though the fines collected covered the cost of replacements for the first several years, as behavior problems increased, so did the inclination to make greater efforts to retrieve the library’s property.

5. Conclusion

This chapter has provided an overview of the motivations behind and definitions of the public library in mid-nineteenth century Boston, as well as some of the ways in which those

⁶⁹ I am all but certain that I saw a description of this policy put into practice in one of the later volumes of the *Proceedings of the Trustees*, which sadly I did not have sufficient time to photograph, and thus cannot reference here.

ideals were put into practice both in terms of physical and intellectual structures and in terms of administrative processes. Out of the four cases to be presented here, the Boston Public Library is exceptional in a number of ways, both historical and analytical.

Historically speaking, Boston is unusual for its intense impact on the shape of the public library movement: library historian Sidney Ditzion has gone so far as to claim that “from the point of view of its influence on every aspect of the American library movement, Boston dominated the field completely” (Ditzion 1947, 5). One indication of this impact is the familiarity of many of the phenomena described above to anyone acquainted with modern public libraries: when Boston did things like allowing the bulk of the collection to circulate or opening its doors to anyone who cared to enter, it was experimenting with completely new ways of running a library, subject to controversy and negotiation. Now, we simply take for granted that this is how public libraries work. Further, Boston is analytically unique among the cases included here because of the relative linearity and cohesiveness of its history. The Boston Public Library was a single institution developed in a single location by a relatively small number of (mostly) men, many of whom already knew each other well prior to embarking on the development of the library, and who had an absolute penchant for writing down everything they did and keeping every word preserved within the halls of the institution they founded. The same cannot be said of any of the other cases.

Indeed, the next chapter will explore a case that could not be very much more dispersed: the Carnegie Library Program, in its earliest development phases, from 1889-1899. The dispersal of Carnegie Libraries is simultaneously what qualifies them for inclusion in this study – the program spread public libraries as buildings and as a taken-for-granted concept all over the United States and beyond – and what makes the case extremely challenging to represent in a similar form to that above, since the primary source material for the program is patchy, ill-preserved, and spread across at least dozens and potentially hundreds of individual institutions throughout the country. Still, it is worth noting here that even the Carnegie program owed a debt to the BPL, both for its role in casting the public library as an important municipal institution, and for sketching the outlines of how such an institution might function, even on a globally remarkable scale.

Case 2: The Carnegie Library Program (~1889-1899)

The development of the Carnegie library program differs from that of the Boston Public Library in many ways. As will be evident from the juxtaposition of these chapters, one of those ways is in the existence and accessibility of primary source data responsive to the research questions posed in this dissertation. Whereas the Boston Public Library grew up as a basically unitary institution with a largely linear history – even housed in the same building since 1895 – the story of the Carnegie library program is a tale that meanders in many directions, through many institutions, and with many loci of control.⁷⁰ This, has strong implications for the task of assembling narrative that adequately describes the program's history. Records related to the Carnegie library program are spread thinly among numerous institutions, including many of the more than 2500 libraries whose buildings Carnegie donated. Though higher concentrations of documents exist in certain places – especially the Carnegie Corporation of New York (CCNY) archives at Columbia University and the Andrew Carnegie Papers at the Library of Congress – a comprehensive inventory of all the places where relevant documents might exist has simply been beyond the scope of this research. Additionally, the strong emphasis placed on efficiency by both Carnegie himself and later the CCNY has further limited the availability and accessibility of extant documents related to the library program: in the 1940s, the CCNY had all of the correspondence related to the library program microfilmed, and the originals destroyed (Bobinski 1969, viii). Undoubtedly this saved a great deal of filing space for the Corporation; however, microfilming was still relatively new at the time, and the microfilmed versions of the letters, now held at Columbia's Butler Library, are generally of quite low visual quality – in many cases actually illegible.

Despite these challenges, however, a relatively complete picture of the motivations, definitions, and implementation of the early years of the Carnegie library program can be pieced together from primary sources obtained on visits to the CCNY Archive at Columbia and to the Library of Congress, and also downloaded from various online repositories, in conjunction with the many broader-scope secondary sources available regarding the program – including Bobinski's *Carnegie Libraries* (1969, written largely out of the microfilmed correspondence now housed at Columbia), Van Slyck's *Free to All* (1995),

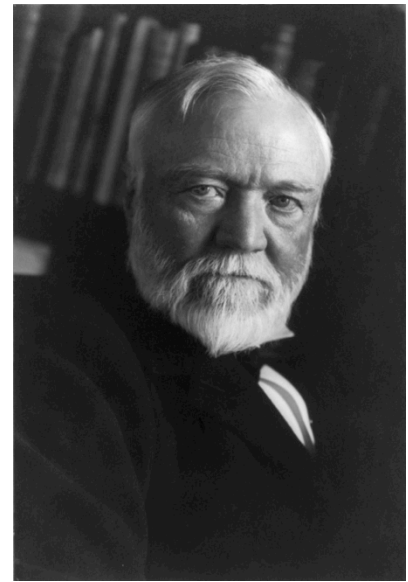


Figure 25: Andrew Carnegie

⁷⁰ One could say that if the Boston Public Library grew up like a sequoia – a large, strong, central support structure off of which branches shoot – the Carnegie library program grew more like English Ivy – weak and convoluted in its connections, but wide-spreading and tenacious nonetheless.

Jones's *Carnegie Libraries Across America* (1997), MacLeod's *Carnegie Libraries in Wisconsin* (1968), and Koch's *A Book of Carnegie Libraries* (1917), among others.

As with the Boston case, this assessment of the Carnegie library program will mainly focus on a limited period of time, encompassing the very earliest, "retail" donations (made selectively, largely based on Carnegie's personal connections to the locations) as well as the first year or so of the "wholesale" period (in which the personal element was progressively withdrawn in favor of systematization and scale), in 1898. However, as in the previous chapter, the analysis here will often reach beyond this date in order to provide further context and to assess how those early decisions played out in practice. In addition, rather than exploring the idiosyncrasies of motivation and implementation across the hundreds of recipient communities, the majority of the focus here will be placed on the central, programmatic features of the Carnegie donations: what motivated the program as a whole, the sorts of users and collections that were envisioned by the donor and early local leadership (the latter as reflected mainly in secondary sources), and the processes through which the program of donating library buildings was pursued in practice.

Before proceeding with this analysis, a brief historical synopsis of the development of the Carnegie library program will help to orient readers to the case.

1. Historical Synopsis

By the time Andrew Carnegie began donating library buildings in the 1880s, the U.S. public library movement was already well underway. Within the 25 years after the 1852 founding of the Boston Public Library, 188 more public libraries were established, and a little more than a decade after that, by 1887, the number had risen to 649 in total, spread across the twenty states that had by then passed library-enabling laws (Bobinski 1969, 7-8). Carnegie's donations would ultimately multiply that figure more than threefold; however, the story of his program of library donations proceeds more as a meandering process of trial and error than an initiative with a clearly-delineated up-front plan (e.g., Bobinski 1969, 187-88). Indeed, despite lasting for more than thirty years, the program never even had a formal name; the term "Carnegie library program," which I use here, is simply one of many monikers applied to it by others in the years since it occurred (Jones 1997, 3).

Carnegie continually underlined the strong link between his own personal history and his reasons for giving libraries in his speeches and writings. As such, understanding a bit about his biography will help to orient the discussion of motivations to follow in the next section here. Carnegie was born in 1835, in Dunfermline, Scotland, the son of a weaver (Hendrick 1932, 10). In 1846, his family fled poor economic conditions in Scotland, emigrating to Allegheny City, Pennsylvania (now part of Pittsburgh), where Carnegie would spend the rest of his youth. Carnegie's formal education ended with his time in Scotland, at age 11, and he began his first job, as a bobbin boy, at age 13 (Bobinski 1969, 9). Around 1850, a local

man, Colonel James Anderson, began to open his personal book collection to the boys and young men of Allegheny, and Carnegie quickly became an avid user of that collection. As Jones notes, since Allegheny lacked a true public library, “books were hard to come by, and Carnegie often credited Anderson for ‘opening...the intellectual wealth of the world’” to him (Jones 1997, 5). And in fact, most of Carnegie’s explicit rationales for his library philanthropy include – or simply consist of – references to his enormous enjoyment of Anderson’s library (e.g., Carnegie 1889b, “Mr. Andrew Carnegie in Aberdeen” 1892, Carnegie Public Library and Carnegie 1892, Carnegie 1898a, Carnegie 1920). As a young man, Carnegie worked his way up in the telegraph and then railroad industry through a mix of self-education, hard work, strong family support, and serendipitous professional connections, ultimately becoming the enormously wealthy steel manufacturer as which he is best known. And although as Van Slyck suggests, he tended to selectively exaggerate or omit pieces of his rise from hardship to prosperity for rhetorical effect in describing his philanthropic philosophy, his early and intense interest in philanthropy is a matter of record. His first recorded expression of a desire to dedicate his surplus wealth to charitable ends appeared at age 33, he gave his first library building at age 46, and he published two enormously influential articles on the value and aims of philanthropy – “Wealth” and “The Best Fields for Philanthropy” – at age 54 (Carnegie 1889a, Carnegie 1889c, Van Slyck 1995, 9). At the time, it was far more typical to leave any thought of charitable donation much later in life, and simply include philanthropic bequests in one’s will (e.g., Jones 1997, 6) – as such, Carnegie’s youthful commitment to this type of activity is notable in its broad deviation from standard practice.

Carnegie donated his first library to his hometown of Dunfermline, Scotland, in 1881 (Learned 1924, 69). In that same year, Carnegie also offered a library to Pittsburgh, the home base of his industrial empire; however, the city was legally unable to accept the library until after 1887, when a law was passed to allow for its maintenance. Because Pittsburgh could not initially accept, nearby Allegheny received the first Carnegie library donation in the United States, in 1886 (Jones 1997, 7-8). Between 1886 and 1897 – which Carnegie would later dub the “retail” period – Carnegie gave more than \$1.8 million for 14 library buildings in six U.S. cities (Bobinski 1969, 13).⁷¹ Five of these cities were in Pennsylvania, in places where Carnegie had business interests (Allegheny, Braddock, Pittsburgh, Johnstown, and Homestead), and each of these received considerably more than a library. As Bobinski suggests,

The underlying idea behind these gifts was clearly that of a general community center and library in which art exhibition halls, lecture and recital rooms, organs, and even gymnasiums and swimming pools were available in addition to

⁷¹ This equates to \$45.9-\$49.6 million in 2013 dollars, depending on which part of the time period one calculates from (1886 or 1897).

reading rooms and bookstacks. Furthermore, all these institutions were given endowments which were certainly contrary to library-giving in the [later, more systematic] "wholesale" period (1969, 13).

The sixth U.S. library donation given in this period was a much smaller gift of \$30,000, to Fairfield, Iowa, in 1892; Jones refers to this as Carnegie's first "typical" library donation (1997, 2).⁷² During the retail period, Carnegie acted largely autonomously, and sometimes in partnership with his wife, Louise (e.g., Carnegie Public Library and Carnegie 1892, Bobinski 1969, 32). And although he did place conditions on these early grants (on which more in a moment), those requirements lacked the more systematic character that the conditions on his later donations would assume.

After 1897 began Carnegie's so-called "wholesale" period. The early gifts in Pennsylvania and Iowa had attracted national attention, and with it further requests for library buildings, which Carnegie generally granted. And, as Bobinski notes, although "Carnegie had no master design in mind when he began giving gifts of libraries...Gradually a plan did evolve" (1969, 187-88). As with the earlier donations, communities during the wholesale period were required to provide a site and commit to annual tax support. However, for these later donations, the minimum tax support requirement was pinned at 10 percent of Carnegie's donation. And the amount of each donation, in turn, came to be determined based on the population of the community - initially self-reported by the communities, but later derived from census figures, which were less likely to exaggerate. Generally, the amount donated was limited to about \$2-3 per capita in the later period. In addition, ever-increasing administrative requirements were imposed, including requiring that requests come directly from elected city leadership (and not the local women's club, newspaper editor, or chamber of commerce), and mandating that building plans be submitted for review prior to construction (Van Slyck 1995, 22, Jones 1997, 26). These procedural shifts will be discussed in more detail in section 4.1, below.

Many - perhaps most - of the program's standard procedures were devised and implemented by the man Carnegie entrusted with the day-to-day management of his library philanthropy after 1897, his personal secretary, James Bertram (e.g., Bobinski 1969, 47, Van Slyck 1995, 22, Jones 1997, 21-25). Bertram shared Carnegie's zeal for efficiency and abhorrence of waste, going so far to adopt Melvil Dewey's simplified spelling along with his employer; this spelling appears in a great deal of his library correspondence, as well as his "Notes on Library Bilding" (Bertram 1911).⁷³ And because of the clear rules that had

⁷² In addition to the U.S. donations, Carnegie gave several more libraries in Scotland during this period as well (e.g., Carnegie 1891, "Mr. Andrew Carnegie in Aberdeen" 1892, Carnegie Public Library and Carnegie 1892).

⁷³ I find Bertram to be a rather fascinating character. From his library correspondence, he often seems brusque to the point of rudeness, and entirely unwilling to suffer fools (a group he seems to have defined quite inclusively), but also extremely capable in dealing with what must have been enormous responsibilities. Bobinski describes him further,

been laid out for the program, Carnegie could essentially delegate the entire process to Bertram, only dipping back into it to spot-check his secretary's work. Carnegie's biographer, Burton Hendrick, recounts Bertram's narration of how his interactions with Carnegie regarding the library program tended to proceed:

"Well, Mr. Carnegie, here are forty or fifty more libraries" [Bertram] would say.

"Are you satisfied they are all right?" [Carnegie] would ask.

"Yes."

"All right, my boy, go ahead with them" - perhaps asking one or two questions to test Bertram's knowledge of the different applications" (Hendrick 1927).

During the wholesale period, the personal dimension that had characterized Carnegie's "retail" donations was thus for the most part excised. So long as a community met the conditions - which, as noted, became a more arduous task over time - they would be given a donation for a building. And it was largely because of this depersonalization and systematization, borrowed from the corporate and engineering worlds in which Carnegie had excelled, that the program was able to take on its ultimately massive scale.

Figure 26, below, illustrates the general shape of the retail period vs. the wholesale period. The red line (and left axis) indicates the average grant amount; the blue bars (and right axis) indicate the number of libraries donated.⁷⁴ It makes clear the extent to which the retail period comprised a small number of relatively large donations - six grants averaging just over \$400,000 each - while the wholesale period consisted of an enormous number of relatively small donations - 1421 grants averaging just under \$27,000 each. It is especially striking to consider that the wholesale average actually includes the largest Carnegie library grant, \$5.2 million given to New York City in 1899, as well as six more of the top ten largest

drawing on his contemporaries' accounts: "To Robert M. Lester, another contemporary Carnegie Corporation employee, 'Bertram was a devoted, meticulous Scot and made a religion out of the Carnegie spirit of giving.' Frank P. Hill described him as methodical, systematic, and a stickler for precedent. Bertram believed in going slowly and being sure of his ground. Brevity was his strong trait. He never used a paragraph when a sentence would do, and a word often served to take the place of a sentence" (1969, 26-27).

⁷⁴ This graph was constructed by compiling all of the state-by-state tables of Carnegie Library grants available on Wikipedia (https://en.wikipedia.org/wiki/List_of_Carnegie_libraries_in_the_United_States). Though that may seem a somewhat questionable data source, I would argue for its sufficiency for this purpose for three reasons. First, the references given on these Wikipedia pages (Miller 1943, Anderson 1963, Bobinski 1969, Jones 1997) are precisely the resources I myself would have chosen to assemble such a list, and Wikipedia's have the advantage of already being in digital, tabular form. Second, in spot-checking the data in the tables pulled from Wikipedia against these sources, I have found no errors (though discrepancies do exist among the sources, and in those cases the most recent source is privileged on Wikipedia). And finally, for the purpose of making a chart such as this one, where the main point is to illustrate the overarching trend, it simply will not matter if a few figures are off here and there - particularly since, having examined several compilations of this same data, I can attest that they cannot be off by much.

grants: St. Louis (\$1 million, 1901), San Francisco (\$750,000, 1901), Detroit (\$750,000, 1901), Washington, DC (\$682,000, 1899), Cleveland (\$590,000, 1903), and Baltimore (\$500,000, 1906). These very large donations, however, are markedly outweighed by the huge number of very small donations during the same period: nearly half (706) of the wholesale period library grants were \$10,000 or less, and just under 85% (1199) were \$25,000 or less (Miller 1943, Anderson 1963, Bobinski 1969, Jones 1997).

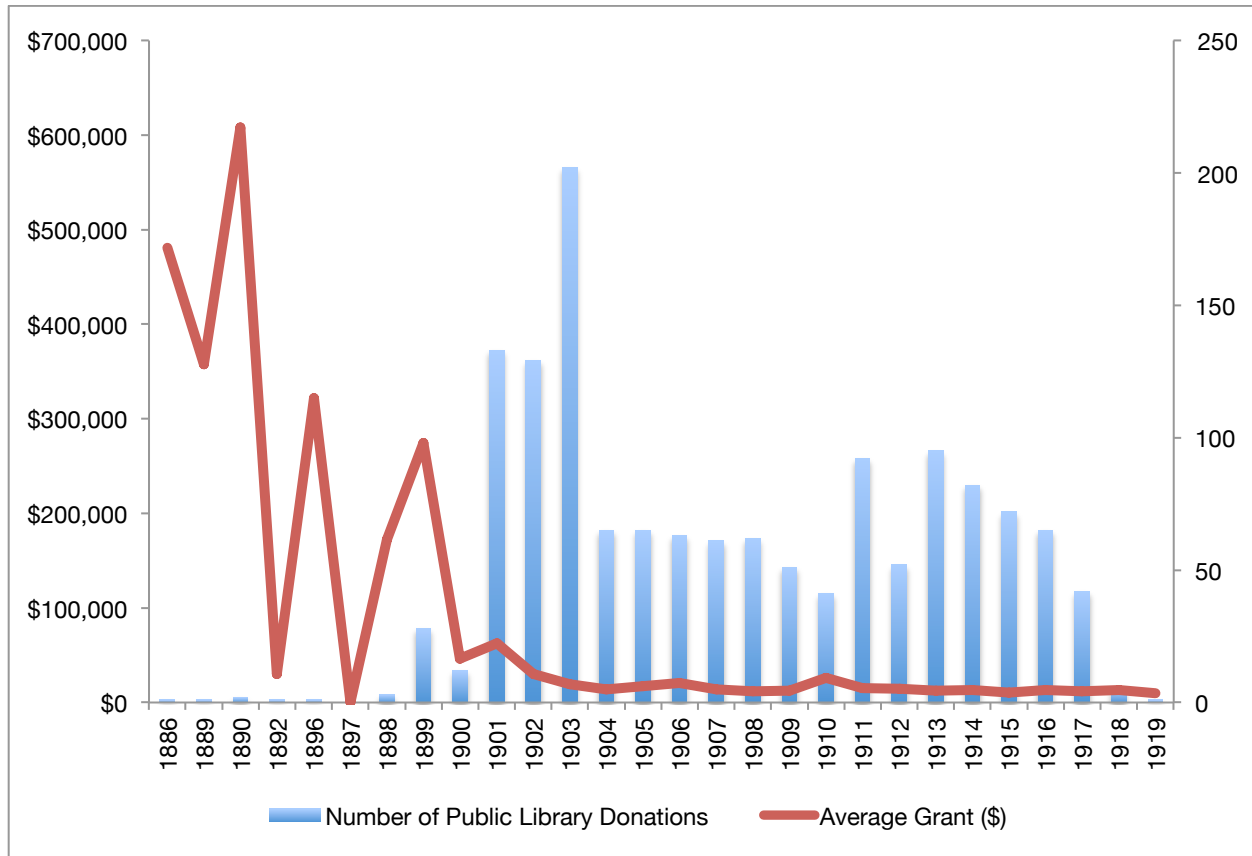


Figure 26: Carnegie Public Library Grants By Year: Number of Donations and Average Grant Amount

Administratively, the Carnegie library program can be said to have passed through three major phases. In the first (1886-1897), it was wholly under the control of Carnegie himself. Conditions applied to these grants, but they were not imposed evenly or stringently. In the second period (1898-1911), the program was almost entirely delegated to James Bertram, with the aid of just one assistant (Macleod 1968, 24, Bobinski 1969, 157). And finally, in the third period the program was administered by the Carnegie Corporation of New York (CCNY), established in 1911 with a mission

to promote the advancement and diffusion of knowledge and understanding among the people of the United States, by aiding technical schools, institutions of higher learning, libraries, scientific research, hero funds, useful publications, and by such other agencies and means as shall from time to time be found appropriate therefor ("Act of Incorporation" 1911).

The world's first modern foundation – called a “corporation” because the word “foundation” had not yet taken on its modern meaning (Van Slyck 1991, 370) – the CCNY took over responsibility for managing the vast majority of Carnegie's charitable work, including the library donations. Of course, the CCNY's initial executive committee consisted of Carnegie, Bertram, and Robert A. Franks (Carnegie's former private financial secretary, turned CCNY treasurer) with loose annual oversight by a board of trustees – and as such, Bertram retained a great deal of his existing control over the library program during the CCNY phase, in his new role as secretary of the Corporation (Bobinski 1969, 26).

The Carnegie library program was placed on indefinite hold in November of 1917, and no new grants were offered after that point.⁷⁵ Officially, this decision was made due to wartime shortages of labor and supplies.⁷⁶ However, the true rationale was likely a bit more complex. For years, questions had intermittently arisen about the effectiveness of the program on the ground, but since all donation-related business was conducted by mail, with no in-person visits to communities, it was difficult to adequately assess its impact (Van Slyck 1995, 217, Jones 1997, 100). Thus, in 1915, the CCNY trustees commissioned an economist, Alvin S. Johnson, to travel around the country visiting Carnegie libraries, and to prepare a report on “the results of the wide provision of Public Library Buildings by Mr. Carnegie and [the] Corporation, with such recommendations as his study may lead him to propose” (Johnson 1916, title page). Johnson did so, and his report is a masterful – and tactful – assessment of the factors relevant to public library success or failure, and the relative contributions of the Carnegie program to those factors, for good or ill. In particular, he notes a high level of non-compliance with grant conditions, as well as significant shortfalls in library startup funding in communities receiving small donations (\$20,000 or less). As Johnson explains, a \$10,000 grant required the community to provide \$1,000 per year for annual maintenance (and communities rarely exceeded the requirement) – but \$1,000, even in 1900, was not sufficient to acquire an initial collection of books while at the same time covering staff pay, utility bills, and other expenses. Thus, these small libraries would undergo a stagnant period just after opening, where they had a building and perhaps some staff, but few books to lend, from it was very difficult to emerge (Johnson 1916, 51-52). Throughout his report, Johnson also emphasizes the importance of skilled librarians, going so far as to aver that “[the] value of a library depends largely upon the character of its personnel” (Johnson 1916, 18), and suggests that many of the donated libraries had inadequacies in this area as well.

⁷⁵ Though six more donations were made in 1918-1919, these had already been under consideration prior to the suspension of the program (Jones 1997, 102).

⁷⁶ Van Slyck suggests that this rationale may have been given in part to spare Carnegie's feelings, as he was reaching the end of his life (1995, 217). This is further supported by the fact that Johnson's report was not made public until after Carnegie's death in 1919 (Van Slyck 1995, 249, note 1).

Johnson's ultimate recommendations – that field agents be deployed to assess communities' likelihood of success prior to committing funds, that the CCNY should provide (and pay the salary of) a skilled librarian to help get each library up and running, and that the CCNY ought to consider supporting the training of professional librarians in addition to donating buildings (Johnson 1916, 57-68) – were received somewhat ambivalently by the Corporation's trustees. James Bertram, in particular, was horrified, and declared that Johnson's proposals "[flew] in the face of Mr. Carnegie's intentions" and introduced precisely "the waste of centralized bureaucratic control," which Carnegie had sought to avoid (cited in Jones 1997, 101). Still, almost exactly a year after Johnson submitted his report, the program was suspended, never to resume. And in 1925, after Carnegie's death, the Corporation began to act on Johnson's recommendation to turn funding in the direction of librarian training,⁷⁷ initiating a program of grants for librarian education, academic libraries, and library professional associations like the American Library Association (Van Slyck 1995, 217).

2. Motivations

The network of motivations underlying the Carnegie library program has a quite different shape from that which lay behind Boston Public. Whereas that earlier institution developed largely through negotiations among a dozen or so co-located, evenly matched, and generally like-minded elite men, the initiation of the Carnegie program depended centrally on Carnegie alone – his philosophies and his objectives – while later phases proceeded through long-distance, asymmetrical, written consultations between James Bertram and representatives from recipient communities. Because the focus in this dissertation is on the initial phases of these projects, its examination of the motivations behind the Carnegie program will focus mainly on Carnegie's own motivations – his motivations for philanthropy in general, and also his motivations for giving libraries in particular – but will also briefly touch on the recipient communities' motivations for making library requests.

2.1 Carnegie's Motivations for Philanthropy

As already noted, Carnegie came to his dedication to philanthropy relatively early in life, and always pointed to his own semi-mythologized rags-to-riches story as an element of his rationale for that dedication. Still, as outlined in "Wealth" and "The Best Fields for Philanthropy," Carnegie's philosophy of philanthropy went beyond a mere extension of personal biography. Three elements of this philosophy – his conception of the obligations that accompany wealth, his desire to help only those willing to help themselves, and his support for evolutionary (not revolutionary) social change – help to lay the groundwork for understanding the extent and particular character of his library donations.

⁷⁷ A suggestion which had since been reemphasized by a second Corporation-commissioned report, by Charles C. Williamson (1923).

2.1.1 Wealth as Virtue, Virtue as Duty

In his thinking about class mobility, Carnegie was essentially a social Darwinist.⁷⁸ In “Wealth,” for example, he suggests that

whether the law [of competition] be benign or not, we must say of it...It is here; we cannot evade it; no substitutes for it have been found; and while the law may be sometimes hard for the individual, it is best for the race, because it insures the survival of the fittest in every department. We accept and welcome, therefore, as conditions to which we must accommodate ourselves, great inequality of environment, the concentration of business, industrial and commercial, in the hands of a few, and the law of competition between these, as being not only beneficial, but essential for the future progress of the race (1889c, 655).

In essence, Carnegie believed that the accumulation of wealth was a natural consequence of possessing superior aptitudes for organizing and managing people and resources, and thus that excess wealth served as a marker of talent and worth. He even goes so far as to call individuals accruing such surplus wealth “the highest type of man” (1889c, 657).

And Carnegie had very particular ideas about what these men (and they were, in general, *men*) should do with this surplus: that is, they should give it away during their lifetime, and not hoard it to be dispersed upon their death (1889c, 658-59). After all, who better to judge the best uses of surplus wealth than individuals who have already proven their moral worth and good judgment by earning it? And how better to ensure that those judgments are executed correctly than to supervise that execution personally? To the extent that millionaires would adopt such a program of active philanthropy during their lives, Carnegie predicted,

we shall have an ideal state, in which the surplus wealth of the few will become, in the best sense, the property of the many, because administered for the common good, and this wealth, passing through the hands of the few, can be made a much more potent force for the elevation of our race than if it had been distributed in small sums to the people themselves. Even the poorest can be made to see...that great sums gathered by some of their fellow-citizens and spent for public purposes, from which the masses reap the principal benefit, are more valuable to them than if scattered among them through the course of many years in trifling amounts (1889c, 660).

Carnegie’s belief in the necessity to distribute surplus wealth during the donor’s lifetime was thus a classic example of *noblesse oblige* for the capitalistic nobility of the industrial age;

⁷⁸ Indeed, Carnegie seems to have been personally acquainted with the originator of social Darwinism, Herbert Spencer. In Carnegie’s dedication of the Pittsburgh Carnegie Institute, for example, he describes receiving “a letter from Herbert Spencer, who had visited Pittsburg with me just after the library was refused. He was bitter about some letters from correspondents in the papers, who explained, to their own satisfaction, no doubt, that my aim was only to erect a monument for myself” (quoted in Koch 1917, 132). Though Carnegie disagreed with Spencer’s assessment in that instance, he also mentions Spencer in favorable terms in “Wealth” (1889c, 663) and goes so far as to label himself a “disciple” of Spencer in his autobiography (1920, chapter XXV).

a social obligation imposed by the mere fact of possessing wealth.⁷⁹ The duty of the wealthy man, to Carnegie, was to first avoid ostentation and provide modestly for dependents, and then beyond that,

to consider all surplus revenues which come to him simply as trust funds, which he is called upon to administer, and strictly bound as a matter of duty to administer in the manner which, in his judgment, is best calculated to produce the most beneficial results for the community – the man of wealth thus becoming the mere agent and trustee for his poorer brethren, bringing to their service his superior wisdom, experience, and ability to administer, doing for them better than they would or could do for themselves (1889c, 661-62).

Thus, “the millionaire [would] be but a trustee for the poor,” duty-bound to expend his wealth on projects which help individuals to improve their own situation (Carnegie 1889c, 664).⁸⁰ And while Carnegie saw many ways of doing this (laid out in “The Best Fields for Philanthropy” (1889a)), libraries were the first major way in which he himself went about fulfilling his own self-imposed obligations.⁸¹

2.1.2 “Nothing for Nothing”

Still, Carnegie did not approve of just *any* kind of charity. In fact, he *disapproved* of most. He was fond of proclaiming, for example, that it would be “better for mankind that the millions of the rich were thrown into the sea than so spent as to encourage the slothful, the drunken, the unworthy,” and that, of every \$1000 given to charity, “it is probable that \$950 is unwisely spent; so spent, indeed, as to produce the very evils which it proposes to mitigate or cure” (e.g., Carnegie 1889c, 662). To avoid these perils, Carnegie favored the use of philanthropy only on projects that would exclusively benefit those who were willing to make their own independent effort to improve themselves. In fact, he goes so far as to

⁷⁹ Or, if you prefer, a much earlier iteration of the *Spider-man* axiom: “With great power comes great responsibility.”

⁸⁰ It is worth noting, however, that there was at least one significant tension in Carnegie’s philosophy of wealth: that is, even as he proclaimed the superiority of the wealthy man, he also frequently romanticized poverty (and especially being born poor). For example, at the dedication of the Braddock Carnegie library, he averred that “The great inventions, the improvements, the discoveries in science, the great works in literature have sprung from the ranks of the poor. You can scarcely name a great invention, or a great discovery, you can scarcely name a great picture, or a great statue, a great song or a great story, nor anything great that has not been the product of men who started like yourselves to earn an honest living by honest work” (Carnegie 1889b, 28). For Carnegie, there seems to have been a hierarchy of esteem, in which the self-made millionaire occupied the top rung, the aspiring poor the next, and the idle – whether rich or poor – the lowest.

⁸¹ Though libraries were the first focus of Carnegie’s philanthropic activities, they were far from the last. Alongside his library donations in the late nineteenth century, Carnegie also donated many church organs, hoping to share and spread his love of classical music. And after the turn of the century, when he handed the reins of the library program over to Bertram, and, perhaps more significantly, received a fully liquid \$400 million as a buyout payment from J.P. Morgan for his interest in Carnegie Steel (Nasaw 2006, 584-86), his donations broadened in scope. These ranged from providing Scottish university students with free tuition to various U.S. educational initiatives and institutions to providing an initial seed fund of \$10 million to found the massive endowment for international peace that still bears his name (Nasaw 2006, xiii, Carnegie Endowment for International Peace 2010).

suggest that the pure fact of *needing* help may well be a sign that one is not *worthy of being* helped:

Those worthy of assistance, except in rare cases, seldom require assistance. The really valuable men of the race never do, except in cases of accident or sudden change. Every one has, of course, cases of individuals brought to his own knowledge where temporary assistance can do genuine good, and these he will not overlook. But the amount which can be wisely given by the individual for individuals is necessarily limited by his lack of knowledge of the circumstances connected with each. He is the only true reformer who is as careful and as anxious not to aid the unworthy as he is to aid the worthy, and, perhaps, even more so, for in alms-giving more injury is probably done by rewarding vice than by relieving virtue (Carnegie 1889c, 663).

Among the few types of causes that met with Carnegie's approval in this regard were "parks, and means of recreation, by which men are helped in body and mind; works of art, certain to give pleasure and improve the public taste, and public institutions of various kinds, which will improve the general condition of the people" (Carnegie 1889c, 663).

In describing the positive features of these types of project, Carnegie seemed especially fond of their tendency to "give nothing for nothing;" to reward users in proportion to their own effort, and thus to exclude the indolent and undeserving from reaping any benefits. Indeed, Carnegie used that exact phrase – "nothing for nothing" – on numerous occasions, including at least four library dedication speeches and his autobiography (1891, 1892, 1895, "Dedication of Washington (DC)" 1903, 1920) – and it truly seems to encapsulate his philosophy on philanthropy. For example, at the dedication of the library he donated at Pittsburgh, he declared that,

What we must seek...for surplus wealth, if we are to work genuine good, are uses which give nothing for nothing, which require co-operation, self-help, and which, by no possibility, can tend to sap the spirit of manly independence, which is the only sure foundation upon which the steady improvement of our race can be built. We were soon led to see in the Free Library an institution which fulfilled these conditions, and which must work only for good and never for evil. It gives nothing for nothing (1895).

This philosophy was both a key motivator for the Carnegie library program in general and, as will be described later in this chapter, a significant factor in casting the particular shape of the program – funding for buildings only, with communities expected to independently construct, fill, staff, and maintain them.

2.1.3 Evolution, not Revolution

A final element of Carnegie's philanthropic philosophy worth mentioning here is its basic conservatism: through philanthropy, Carnegie hoped to promote evolutionary social change, rather than sharp disruptions in the status quo. This disposition is evident

throughout his writings and speeches. As he suggests at an 1891 library dedication, “The very first grain of knowledge which men must acquire in their advance is, that success in life is not to be gained by violence, or by agitation; not by revolution, but by evolution, and by strenuous, unceasing, labour and thrift” (1891, 34-35). And elsewhere, he advises that change ought to be gradual, like the re-grading of the city of Chicago, which had been “raised twelve feet by...slow, sure, but steady and unceasing progress” – in a similar way, he avers, “will human society be surely most elevated” (“Mr. Andrew Carnegie in Aberdeen” 1892, 16-17).

This promotion of gradual change over disruptive rifts makes a great deal of sense considering Carnegie’s position and era. The late nineteenth century saw the rise of great labor unrest, including within Carnegie’s own steel empire. In particular, the violent 1892 strike at Carnegie’s Homestead Steel Works had a significant personal impact on Carnegie, even though he tended to publicly minimize the role of his own management and of systemic class inequalities by blaming the strike on low-caliber legacy employees already there when he purchased the plant (Van Slyck 1995, 102). And in parallel to these phenomena, the American socialist movement was also on the rise, supporting the labor uprisings and generally seeking to empower the working class vis-à-vis their employers. Given Carnegie’s position in society – that is, at the apex of wealth and at the head of a major industrial organization – his vehemence in opposing revolutionary change of the sort demanded by the socialists is unsurprising. Still, it does provide further insight into why he chose to promote the philanthropic targets that he did. Parks, libraries, musical instruments, even education – all of these can be seen as having a potentially calming influence on their users, as well as providing some element of “bread and circus:” healthy forms of recreation and intellectual engagement to help distract the worker from his tenuous or oppressive economic situation. As Carnegie himself declares in defending the place of fiction in public libraries (and specifically in the library donated near his steel works in Braddock, PA):

When exhausted in mind and body, and especially in mind, nothing is so beneficial to [workers] as to read a good novel. It is no disparagement of free libraries that most of the works read are works of fiction. On the contrary, it is doubtful if any other form of literature would so well serve the important end of lifting hard working men out of the prosaic and routine duties of life (1889b).

It would be overly cynical and narrow-minded to suggest that Carnegie’s philanthropy was *primarily* an elitist attempt to maintain the existing hierarchy among social classes;⁸² however, this “defense mechanism factor,” as Ditzion calls it (1947, 156), was certainly present in his reasoning.

⁸² As some have (Harris 1975b).

2.2 *Carnegie's Motivations for Giving Libraries*

Carnegie's support for libraries as the earliest outlet for his philanthropic tendencies was to some extent simply a logical extension of the philanthropic philosophy outlined above. However, Carnegie's attachment to libraries in particular was neither purely philosophical nor fully consistent over time. This section will articulate the linkage between Carnegie's philanthropic philosophy and his library donations, examine the role of his personal history and tastes, and finally note some of the ways in which Carnegie's motivations for library donations shifted over the course of the program – specifically, from an early paternalistic, personally-invested mode to a later depersonalized, systematic mode.

2.2.1 Providing a Ladder for the Aspiring

Carnegie often made explicit the close connection between his “nothing for nothing” philanthropic philosophy and his predilection for giving libraries. In several instances, he drew this tie using the metaphor of the library as a ladder for the aspiring to ascend, in both thinking and way of life. For example, in his dedication of the Pittsburgh Carnegie library, Carnegie proclaims:

There is nothing here that can tend to pauperize, for there is neither trace nor taint of charity; nothing which will help any man who does not help himself; nothing is given here for nothing. But there are ladders provided upon which the aspiring may climb to the enjoyment of the beautiful and the delights of harmony, whence comes sensibility and refinement; to the sources of knowledge, from which spring wisdom; and to wider and grander views of human life, from whence comes the elevation of man (Carnegie 1895).

On the individual level, Carnegie saw libraries (along with museums, parks, and music halls) as objects of philanthropy that would by their very nature benefit only the “deserving:” those willing and able to make an effort to help themselves independently of any assistance received.

Further, Carnegie's desire to help only those willing and able to help themselves extended to *communities* as well as individuals – a fact which had a profound influence on the shape of the library program as a whole. Only communities who pledged to provide a building site and perpetual tax support for a library would be given money for a structure. After all, how else would a community gain a sense of ownership in their library, if not by directly contributing to its establishment and maintenance? Carnegie's thinking on this topic was fully formed well before his library donations ramped up to their eventual enormous scale. Again, at Pittsburgh:

When this library is supported by the community, as Pittsburgh is wisely to support her library, all taint of charity is dispelled. Every citizen of Pittsburgh, even the very humblest, now walks into this, his own library, for the poorest laborer contributes his mite indirectly to its support. The man who enters a library is in the best society this world affords; the good and the great welcome

him, surround him, and humbly ask to be allowed to become his servants; and if he himself, from his own earnings, contributes to its support, he is more of a man than before (1895).

Just as no individual could gain the full benefit of a library's books without putting in the effort to read them, so Carnegie believed that no community could gain the full benefit of receiving a library grant without exerting its own independent initiative to work out the details of establishing the actual institution. And although time would show that in fact, more guidance might have been preferable at both the individual and community levels (Johnson 1916, Macleod 1968), Carnegie never backed down from either of these convictions: in his autobiography, written late in his life, he reiterates that "The fundamental advantage of a library is that it gives nothing for nothing. Youths must acquire knowledge themselves. There is no escape from this" (1920).

2.2.2 Personal History and Affection for Books

Still, as noted, Carnegie's decision to give libraries was not based solely – or even necessarily primarily – on cold, dispassionate social philosophies. As he admits in the second of his two influential early articles on wealth, "It is, no doubt, possible that my own personal experience may have led me to value a free library beyond all other forms of beneficence" (Carnegie 1889a, 689). And from there he launches into the story of Colonel Anderson's library – which, though already described above, will be helpful to include in his own words, as he repeated these exact phrases so many times, with so little variation:

When I was a boy in Pittsburg, Colonel Anderson, of Allegheny, – a name I can never speak without feelings of devotional gratitude, – opened his little library of four hundred books to boys. Every Saturday afternoon he was in attendance himself at his house to exchange books. No one but he who has felt it can know the intense longing with which the arrival of Saturday was awaited, that a new book might be had. My brother and Mr. Phipps, who have been my principal business partners through life, shared with me Colonel Anderson's precious generosity, and it was when revelling in these treasures that I resolved, if ever wealth came to me, that it should be used to establish free libraries, that other poor boys might receive opportunities similar to those for which we were indebted to that noble man (Carnegie 1889a, 689).

A love for literature and reading was thus instilled in Carnegie from an early age. And although Carnegie placed a stronger public emphasis on his own explorations in Colonel Anderson's library in fostering this interest, his parents' influence also seems to have played a role. Carnegie's father, together with a few other local weavers, had helped to organize the first proto-library in their Scottish home of Dunfermline (Bobinski 1969, 11-12, Jones 1997, 12), and his mother was reputedly "an incessant reader and an independent thinker" (Learned 1924, 66); both of these facts helped to solidify Carnegie's love of books and libraries from an early age. And moreover, these early influences and experiences had

the effect of exposing Carnegie to and teaching him to appreciate a much wider breadth of literature than most of his social class. As Koch describes,

[Carnegie's] youth knew the spell and the inspiration of Burns and Shakespeare and those noble old ballads in which the idealism, the passion, and the tragedy of Scottish life found such moving and dramatic expression. Self-made in his independence of material help, Mr. Carnegie was singularly fortunate in the ancestral influences which penetrated and enriched his nature far below the region of his practical activity and efficiency, that deeper part of him which has found expression in these later years, and has asserted its priority of spiritual importance over the executive side of his character (1917, 5).⁸³

Thus, when Carnegie selected libraries as his first favorite locus of philanthropic work, it was at least partially a reflection of the literary influences that pervaded his own upbringing and home life.

Still, most people who love to read obviously do not extend that enthusiasm into a global program of philanthropy. That Carnegie *did* do so shows the influence of, first, the philanthropic philosophy outlined above, but also a second factor: that is, at heart, Carnegie had great faith that most people – or at least most English-speakers – were very much like himself, and given similar opportunities, he believed they could propel themselves to equally great social and economic heights. As Macleod notes,

Carnegie's belligerent faith in America led him to the same beliefs [as had his experiences as a working-class youth in Pittsburgh]; Americans, as a nation, were like himself. The number of tons of paper consumed annually proved that America's literary culture, if not the world's most distinguished, was its most widely diffused (1968, 16).

This conception of the library user as a person with similar enthusiasms and ambitions to Carnegie himself will form a piece of the next major section of this chapter, on imagined users. Here it will be sufficient to note that this kind of assumption undoubtedly played a role in motivating the library program: after all, if a key criterion for "good" philanthropy is

⁸³ Carnegie's enormous love for Burns in particular comes through nowhere more clearly than in his absolutely giddy library dedication speech in Burns's hometown of Ayr. As he writes:

in the heart of every true Scot Ayr occupies a unique place. Not only our own race, but the most advanced and best members of other races, gather here from all parts, as the Mohammedan to Mecca, or the Hindoo to Benares, to find that which is not to be found elsewhere ; that one man of pure genius has become the solvent which melts and forms into one homogeneous element the entire national life of a people, – a poet who is equally the poet of the professor, peer, and peasant; as much the solace, the guide, and inspirer of the noble and the millionaire as he is of the poorest labourer. ... In the different phases of life every class in Scotland draws from the same source the lines or the stanza which cheers or which sheds light upon his path and encourages him; and these selected winged words are the same words, no matter what the wealth or rank, or the poverty, or the condition of the individual. The Scots as a race are in essence one; the differences which divide are shallow; the sentiments which unite are deep, and fuse at the touch of Burns (Carnegie Public Library and Carnegie 1892, 19-20).

that the recipients be deserving, who better to target than people the benefactor perceives as being much like himself?

A final point worth raising in this segment is that the personal dimension of Carnegie's giving – his upbringing and his love of reading – was Carnegie's own preferred narrative by a wide margin, well beyond any tie to his elaborate philanthropic philosophy. Indeed, his stated justifications for library donations very rarely went further than recounting the story of Colonel Anderson and rhapsodizing a bit over the canon of English-language literature (e.g., Macleod 1968, 3-4, 8, 20, 29, Van Slyck 1991, 364). To a certain extent, this vague hand-waving on Carnegie's part demonstrates the degree to which libraries had already become a taken-for-granted, presumed-desirable part of the municipal landscape in the fifty or so years since the founding of the Boston Public Library. As Macleod notes, by Carnegie's era, "A library was...a worthy object of philanthropy, hardly requiring explanation" (1968, 20). Thus, to Carnegie, a brief nod to his childhood love for reading could easily seem more than sufficient as rationale.

2.2.3 Shifting Motivations Over Time

Finally, it is important to note that Carnegie's motivations for giving libraries changed significantly over time, with a particular split between his early "retail" giving and his later "wholesale" period. In the early years, Carnegie acted much more like most other philanthropists, then and now, by making donations to communities with which he had some personal tie: his hometown in Scotland, the Pennsylvania towns in which he grew up and built his steel empire, his favorite poet's birthplace, and so on. His affection for these communities comes through in his communications with them (e.g., "Free Public Library: Johnstown, Pennsylvania" 1890) and also in his dedicatory speeches at the openings of their libraries – a practice which he almost entirely abandoned after 1900 (e.g., 1889b, 1891, "Mr. Andrew Carnegie in Aberdeen" 1892, Carnegie Public Library and Carnegie 1892, 1894, 1895, 1898b). And beyond mere words, his actions also demonstrate his stronger emotional attachment to these locations: many of the municipalities with which Carnegie had a personal tie received much more generous and elaborate donations than might otherwise have been prescribed based on their population size and level of need. For example, some of his early Pennsylvania donations received endowments along with buildings – a practice which Carnegie explicitly opposed, on the grounds that it diminished the community's investment in maintaining the institution, but nonetheless performed several times early on, out of personal sentiment (Bobinski 1969, 76). And these donations also extended well beyond mere library buildings, providing swimming pools, museums, lecture rooms, and gymnasias as well (Learned 1924, 69-70, Jones 1997, 9). As Jones notes, these early donations were quite out of line with the philosophy Carnegie had already expounded in "Wealth:" instead of disseminating surplus wealth based on desert, in these cases he gave enlarged gifts out of "pride and paternalism" (1997, 7).

In his later giving, Carnegie hewed much closer to his stated ideals, explicitly rejecting the notion of personal sentiment as a consideration in donation decisions. This rupture between donor sentiments and philanthropic activity marked a radical innovation in philanthropy, and was a key element in the Carnegie library program's achievement of enormous scale. Though communities would continue to try to demonstrate reasons why Carnegie should care particularly strongly about them,⁸⁴ after about 1898, such pleas became almost entirely irrelevant to whether or not they would be granted a library. Instead, the decisions were made based solely on the community's commitment to fulfill the obligations that went along with the grant – a site for the library plus at least 10% of the donation every year for maintenance. This allowed Carnegie to delegate essentially the entire program to Bertram, who ran it in a strictly impersonal manner: as Bobinski describes, Bertram

judged proposals strictly on their merit. Personal relations or considerations never influenced his judgment. No worthy applicant was to be rejected, and, yet, no unworthy one was to be accepted. As a result, some felt that Bertram was inconsiderate, austere, legalistic, and overly critical, but those who knew him well characterized him as sympathetic, sincere, kindly, logical, practical, fair-minded, and socially affable. He was a hard-working secretary who did his job to the best of his ability (1969, 30).

Under Bertram, the motivations behind each donation became identical, and purely in line with Carnegie's businesslike philosophy of giving. Only the deserving would benefit, and that desert would be proven through the community's written pledge to contribute land and tax money to any library donated. Any community that cleared that bar was approved for a Carnegie donation – no personal or emotional calculation necessary.

2.3 Recipient Community Motivations

For the recipient communities, as might be expected, the motivations were widely varied, though they did exhibit certain commonalities. In particular, as in Boston, local pride, education, and restraint from vice all helped to motivate library requests in the Carnegie program. Because the focus here is mainly on the systematic features of the program as a whole, however, I will only briefly touch on each of these three local motivations before moving on.

2.3.1 Local Pride

Whereas in Boston, education stood out as the number one most-cited motivation for founding a public library, in the Carnegie library program, local pride seems to have taken the lead role. With Carnegie library donations being offered liberally all over the country in the early years of the twentieth century, local communities quite reasonably viewed possessing one as a way of keeping up with the neighbors. As Koch writes (during the

⁸⁴ For example by emphasizing their population of Scottish descent, local links to Carnegie Steel, or even their possession of a Scottish city name.

waning years of the program):

the fact that one town has a Carnegie library is an incentive to its less intellectual or less enterprising neighbor to provide equally good library facilities for its citizens; and more than one community has been spurred to action in this matter by seeing what was being done by its rivals (1917, 12).

And Jones concurs, suggesting that this sort of competitiveness

was especially important to young towns trying to distinguish themselves from surrounding communities. A public library on Main Street was brick-and-stone proof of civic superiority, concern for education and high-minded culture, and the commercial vigor to support it (1997, 17).

With Carnegie handing out money for libraries to any community that met his conditions, many communities seem to have simply thought, “well, why *not* have a library?” – without fully considering all the hard work and resources that would be necessary to make that library flourish. As Bobinski notes,

Many communities were totally ignorant about the techniques and problems of running and supporting a public library. Falling under the spell of Carnegie philanthropy, they applied for a building. Here was something for nothing, or, at least, so it seemed (1969, 101).

Many communities would soon discover that building up and maintaining a library was beyond either their economic means or their expertise or both – and many Carnegie-donated libraries would fail as a result – but these practical shortfalls tended not to outweigh the demands of civic pride and competitiveness in the initial push to secure a Carnegie donation.

2.3.2 Education

Whereas in Boston, the connection between libraries and education had to be articulated and argued for, by Carnegie’s era it had become a taken-for-granted element of what libraries did. References to the inevitable educational benefits of a Carnegie library donation pepper the correspondence sent by communities requesting or receiving libraries. For example, as the President of the Carnegie Library of Atlanta wrote to Carnegie in 1899,

That the library is an educational force in a community, second only to the public school, is a well known fact, and the good that will result from such a library... will be the means of making Atlanta the center of literary activity in the Southeast, as it is already the commercial center ("Free Public Library: Atlanta, Georgia", Letter from W.M. Kelley, December 31, 1899).

And in that same year, the Third Annual Report of the Pittsburgh Carnegie Library cites a then-recent National Education Association publication to make a similar point. Strongly echoing sentiments expressed in Boston decades earlier, its content and source illustrates just how mainstream the library’s educational role had by then become:

The community should be led to regard the library as a necessary part of a system of public education, just as essential as the common school. ... A free public library is the adult's common school. Pupils should know what a library is, what it contains and how to use it....The school trains for a few years, the library for a lifetime (Carnegie Library of Pittsburgh 1899, citing the National Education Association).

Still, there was one vital difference between the conception of library-based education among Boston's early trustees and that in many of the Carnegie library recipient communities: the inclusion of children. As Van Slyck notes,

whereas earlier libraries had been exclusively adult affairs, only the smallest of the new libraries failed to provide a special reading room for the use of children. Young readers found in the children's reading room a portion of the public landscape that catered directly to their needs (1995, 41-42).

Indeed, by 1900, such children's rooms were seen as an absolute necessity in any new library (Van Slyck 1991, 371) – a radical shift in the fifty years since the BPL's founding.

2.3.3 Social Palliative

Finally, leaders of Carnegie recipient communities, like the Bostonians before them, expressed a belief that the library would be a boon to local morals, as it would provide a healthy alternative to saloons and vice. For example, a local library donor in Allegheny averred that the library must be kept open on Sunday evenings, because at that time "the temptations are the most dangerous to young men, and time hangs the most heavily on their hands" ("First Annual Report, Allegheny" 1891). And as Jones suggests, such sentiments were not uncommon:

[towns'] spokespeople fretfully relayed information that their young, single men, particularly those newly transplanted from rural areas, were prone to drinking. They believed the best remedy for the protection of both the men and the town would be a public library (1997, 19).

At the turn of the twentieth century, most of the regions receiving Carnegie libraries – the Midwest, the Great Plains, the West coast – were only just moving beyond their rowdy frontier beginnings. In such locations, the library was seen as a civilizing influence; an institution that would simultaneously modernize the community, educate the populace, and reduce antisocial behavior.

And, returning to Carnegie himself for a moment: though the desire to restrain from vice and build moral character appears only rarely in his written discussions of the library program, the main context in which it does so is telling. That is, the most significant instance of Carnegie painting the library as a purveyor of virtue is in his dedication of the Carnegie Library at Homestead. As noted earlier, Homestead's Carnegie library was donated just a few years after a major labor strike had turned deadly and injured many. The

strike was also a focal point of opposition to Carnegie library donations by labor leaders like Eugene Debs (Bobinski 1969, 103). Thus, when Carnegie asserts to those same workers that a library-based workingmen's club "will do more than improve manners, it will insensibly tend to establish a higher code of conduct, a stricter regard for the proprieties of life, and to produce the class of man incapable of anything disgraceful, a sober, self respecting, industrious, educated saving workman" (1898b, 6), it takes on a simultaneously tenser and more paternalistic connotation than it might in another community. Carnegie seems to have adopted the role of concerned city father in this instance, worrying for the moral condition of his unruly Homestead workers just as women's clubs across the country worried for that of their local unruly young men. Of course, unlike those women's clubs, Carnegie undoubtedly hoped that the library's positive effects in improving the "moral condition" of his workers would simultaneously ameliorate the risk of further labor unrest in his corporation, thus indirectly aiding his bottom line – a significant side benefit.

3. Definitions

The foregoing section described the major motivations behind the Carnegie library program, in this case primarily by Carnegie himself, and secondarily by the communities making requests for library buildings. This section, in turn, will assess how these constituencies envisioned the user base and collections for these libraries during the early years of Carnegie library giving. The discussion will further highlight a key difference between this case and Boston Public: that is, whereas the BPL grew up as a largely ideologically consistent whole, the Carnegie program exhibited a stark divide between the central, high-level intentions of program leadership and the diffuse, on-the-ground execution of library donations in communities. As in the previous chapter, the first part of this section will explore what kinds of people the leadership (i.e. Carnegie and Bertram) thought would be using the libraries, and will briefly note some data regarding how many total individuals obtained library access through this program; the second part will discuss some of the challenges in collection building wrought by the program's structure, and will briefly touch on Carnegie's vague and not especially influential vision for what the libraries' collections ought to contain.

3.1 Users

As noted, the discussion of users for the Carnegie program will follow the same structure as that for Boston: first intended composition, then quantities. However, two significant differences from Boston are notable at the outset. First, although Carnegie had quite a clear vision of the types of users he wanted to see use his libraries, as indicated above, that vision was entirely disconnected from practice in all but the very earliest donations. That is, while in places like Pittsburgh, Allegheny, and Carnegie (PA), Carnegie actually did hold some ongoing influence over library practices ("Free Public Library: Pittsburgh, Pennsylvania"

1890, "Free Public Library: Carnegie, Pennsylvania" 1898), for the vast majority of his donations, he entirely disappeared from the scene after the money was disbursed. No oversight, no string-pulling to make sure that the libraries were run in any particular way, or that they were made welcoming to particular groups of people. Thus, the gap between intentions and execution was much greater when it came to Carnegie's own intentions than it was for those of local leadership in each recipient community, who were able to have a much more direct impact. Both will still be discussed, but this fact will be useful to bear in mind.

Second, the number of individual users touched by the Carnegie library program is significantly more difficult to measure than that of the early years of Boston Public, due to the diffuseness of the program. Still, some useful numbers can be gleaned from Learned's analysis of the program in *The American Public Library and the Diffusion of Knowledge* (1924), as well as from Bobinski's canonical *Carnegie Libraries* (1969), and that is what has been done here.

3.1.1 Intended Composition

Given the geographically broad nature of Carnegie's library giving, it is perhaps unsurprising that the resulting libraries had to contend with a much more diverse range of potential users than had been the case in Boston. While many of the issues that emerged in Boston – personal worthiness, building better workers, provision for casual readers, ambivalent gender dynamics – also appeared in the Carnegie program, additional dimensions, especially related to service for children and for African-Americans, became more salient in the latter program. And as noted above, there was often a significant divide between the sorts of users Carnegie himself envisioned for his library donations and the sorts of users that recipient communities actually welcomed through the door. This section will focus mainly on the user-related themes from the Carnegie program that differ from those in Boston: in particular, (1) its particular vision of class mobility, (2) its strong focus on children, and (3) its encounters with the racial dynamics of library use in the post-reconstruction South.⁸⁵

3.1.1.1 *The Aspiring Working Class*

Class was a huge, dominant aspect of Carnegie's personal imaginary of the library user.

⁸⁵ It bears noting at this point that another possible subsection here might have related to the treatment of gender issues, as these were not insignificant. However, to a great extent, as *users*, women were accorded much the same rhetorical treatment and actual usage rights in the Carnegie libraries as they had been in the Boston Public Library (e.g., Carnegie Public Library and Carnegie 1892, 26-27, 1909). That is, they were always explicitly welcome to enter and read, and were often also hired as staff. The major divergence between the two programs with respect to women's roles actually emerges in the area of *administrative processes* (particularly their strong role in garnering local support for libraries, as opposed to their total disenfranchisement in the Carnegie library request process (e.g., Jones 1997, 52)) which will be discussed later in the chapter.

When Carnegie described the people he hoped his donations would benefit, he consistently emphasized aspiring members of the working class above all others – and in the early cases, especially workers in his own factories. Similar to Boston, the issue of personal worthiness loomed large, as did the utilitarian appeal of improving the working class through education. However, where in Boston these two pieces could be teased apart, for Carnegie they were one and the same. As described in the sections above, in Carnegie’s eyes, one proved one’s worth through hard work. Thus, even when Carnegie pronounced hopes that his libraries would benefit the poor, he really meant only the *working* poor – not the truly destitute – since only those willing to work were deserving of his help. Indeed, Carnegie often drew an explicit distinction between what he called the “swimming tenth” and the “submerged tenth,” declaring that only the former was worthy of assistance (e.g., Carnegie Public Library and Carnegie 1892, Carnegie 1895, Carnegie 1909). For example, in one library dedication speech he expounded that “The province of individual benevolence is not the ‘submerged tenth’ but the ‘swimming tenth’ immediately above, who strive to keep their heads above water as self-supporting and self-respecting members of the community” (1909). For Carnegie, the destitute might be worthy of pity, or perhaps assistance from the state (e.g., Carnegie 1889a, 685), but not the aid of philanthropists.

Privileged even above other sorts of working-class strivers in Carnegie’s eyes, moreover, were his own employees. Most of the earliest U.S. Carnegie library donations were quite explicitly built for Carnegie Steel workers, and involved a (less-than-ideologically-consistent) mix of paternalistic benevolence and personal identification with the employees on Carnegie’s part. Carnegie seems to have hoped that these gifts would inspire closer, more familial-feeling relationships between himself and his thousands of employees, in addition to producing a more educated (and thus more useful) workforce for his factories. Carnegie’s dedication speech at the library complex adjacent to his Edgar Thomson Steel Rail Works in Braddock, Pennsylvania makes clear both of these aspirations for his employees as library users – simultaneously beholden to their benefactor as a paternal figure, yet actively pursuing personal betterment and professional advancement. In it, for example, he explains,

It being therefore impossible for the employers of thousands to become acquainted with their men, if we are not to lose all feeling of mutuality between us, the employer must seek their acquaintance through other forms, to express his care for the well-being of those upon whose labor he depends for success, by devoting part of his earnings for institutions like this library....By such means as these we may hope to maintain to some extent the old feeling of kindness, mutual confidence, respect and esteem which formerly distinguished the relations between the employer and his men (1889b, 2).

And later in the speech, he proclaims his faith in the benefits of a more educated workforce, which he claims run counter to the views of many of his peers:

I have heard employers say that there was great danger that the masses of the people might become too well educated to be content in their useful and necessary occupations. It has required an effort upon my part to listen to this doctrine with patience. It is all wrong ; I give it an unqualified contradiction. The trouble between capital and labor is just in proportion to the ignorance of the employer and the ignorance of the employed. The more intelligent the employer the better, and the more intelligent the employed the better. It is never education, it is never knowledge, that produces collision. It is always ignorance on the part of one or the other of the two forces (1889b, 25).

In Braddock, as in other communities linked to Carnegie's Pennsylvania steel empire, the elaborate complexes Carnegie gave, well beyond the modest library buildings that would later become the program's signature, were "unmistakably built for the use of Carnegie employees and their families, 'perks' for his 'company' towns" (Jones 1997, 9).

Beyond this early set of donations, however, as the Carnegie program transitioned into the wholesale period, the overlay of paternalism faded a great deal, leaving behind a more consistent focus on the betterment of the working class as a whole. By 1909, Carnegie's rhetoric of familial relations between workers and employers had largely fallen away, but not his tight focus on workers as library users: as he declared then,

no class in the community is to be benefitted so clearly and so fully as the great mass of the people, the wage-earners, the laborers, the manual toilers. It is for them that the Free Library places within reach the ladders upon which the aspiring can climb....The Free Library is the Library of the working classes (1909).

Indeed, from roughly 1897 onward, Carnegie's library program began to hew much more closely to his previously stated ideology of philanthropy and desert, as expressed in "Wealth" and "The Best Fields for Philanthropy." Where his retail donations exhibit much of the personal sentiment and over-indulgence that he abhorred in those pieces, the Bertram-administered wholesale donations carefully excised all personal feeling, paring the philanthropic process down to a checklist calculus, wherein filling the proper boxes would not only virtually guarantee a donation, but would also produce the amount to be donated through pure mathematics. Still, a particular vision of the user was nonetheless incorporated into this supposedly purely rational calculus: in order to receive a donation, communities had prove their desert by showing that they were willing to work to produce the library themselves. By channeling the request properly through the local officialdom, providing a site for the building, and legally ensuring a tax base for supporting the library proportionate to the amount of the donation, communities showed their initiative and commitment, and thus their merit as deserving recipients – members of the "swimming tenth" rather than the "submerged."

In practice, within recipient communities, the focus on the working class became more

diffuse and variable. In Allegheny City, the 1893 Annual Report suggests that usage there had remained largely in keeping with Carnegie's hopes, reporting that "It may be safely stated that the majority of readers are from the great middle class or working class and the library would therefore seem to be reaching the very people for which it was founded" ("Third Annual Report, Allegheny" 1893, report of William M. Stevenson, Librarian). And in Carnegie, Pennsylvania, the situation seems similar: the secretary of the local library commission writes that "the largest percentage of the people here [patronizing the library] are of the working class and these are the people we would like to get interested" ("Free Public Library: Carnegie, Pennsylvania" 1898, letter from William Shinn to Carnegie, July 26, 1912). Still, it is worth noting that in Allegheny, Carnegie had involved himself quite personally in the administration of the library, while in the town of Carnegie, the library commission was at the time of that letter hoping to procure further funding from Carnegie himself – and thus their depiction of their users may have been colored by a desire to please, and thus extract more money from, the donor.

Further outside of Carnegie's direct line of sight, it seems that the working class was not always welcomed so warmly. As MacLeod describes, Wisconsin Carnegie library boards, generally composed of middle- and upper-class citizens, often had difficulty understanding how to make libraries appealing to working men:

A Wisconsin architect suggested providing rooms 'where men can read and smoke without feeling it necessary to change their working clothes.' He recommended that there be a separate entrance so that they could enter unobtrusively. Accordingly, in 1909 Madison's trustees hired an attendant and opened a reading room in the basement. There men could read magazines or duplicate copies of some of the less demanding books upstairs. In order to avoid the demeaning implications of a basement room, the board advertised it as a cosy spot and invited men to consider it as their private club in town (MacLeod 1968, 90-91).

However, the men did not embrace this "club" with open arms, preferring the local saloons as social spaces – and the reading room was thus shuttered after three years of disuse (MacLeod 1968, 91). And Van Slyck describes how library boards across the country, echoing early attitudes in Boston, developed biases regarding users based on a distinction between "serious readers, or 'scholars,' and casual visitors." The writings of these boards, she explains, reveal that,

the difference between these two groups was not the frequency with which they used the library, but the sort of reading they did there. Serious readers were those who consulted reference materials and read nonfiction books, whether in the arts and humanities or the fields of science and technology, while casual visitors came to the library with 'no particular book or subjects in mind.' By this definition, *serious readers tended to be middle-class city dwellers, searching for ways to enhance their business concerns, or using their leisure time in the pursuit of highbrow*

culture. In contrast, 'casual' reading coincided with working-class reading habits (like reading the daily newspaper in the library), or with middle-class assumptions about working-class reading habits (like a preference for fiction). What was left unsaid was that library boards actively reinforced this class-based hierarchy of readers by commissioning buildings that isolated middle-class users from both the messy realities of library service and the less genteel library users (1995, 98-99, emphasis added).

On the other hand, as Garrison thoroughly explicates in *Apostles of Culture*, these issues were not confined to Carnegie libraries – far from it. In fact, in the decades surrounding the turn of the century, she notes,

When public libraries were established, urban workers used them, to be sure, but not in the numbers for which librarians had hoped. Despite the best intentions of librarians, the subtle library paternalism – with its nineteenth-century formula designed to support and stabilize the status quo – tended to repel many working-class readers (1979, 49).

The alienation of the worker in libraries of this era, far from being an exclusive feature of Carnegie-donated libraries, seems to have been endemic to the period's public libraries in general. The only difference in Carnegie's case is that this alienation went directly against what the donor would have wished – but because beyond the earliest Pennsylvania donations, Carnegie, Bertram, and the CCNY executed no ongoing oversight over donations, there was no way to impress Carnegie's wishes upon local leadership in any meaningful way, and thus little chance of his donations independently bucking the tide of library-based class elitism.

3.1.1.2 *Children: Nuisance and Opportunity*

Carnegie himself had very little to say about children as library users, except to the extent that he expressed the desire to provide reading materials for “other poor boys” like himself (Carnegie 1889a). Yet, by the late nineteenth century, accommodating children in public libraries was a fast-growing trend. As E. H. Anderson, librarian of the Pittsburgh Carnegie Library, notes in his First Annual Report,

The idea of a separate children's department has so taken hold of the library world that no public library can any longer claim to be up with the times without this important feature. The libraries of Boston, Buffalo, Brooklyn, Denver and other cities have recently provided such departments, and their popularity and usefulness have been fully demonstrated. No part of a public library's work brings bigger returns. If we could put all our juvenile books in one large room, where the children could get at them under proper supervision, and where they could also be issued for home use, we should receive the thanks of the children and older people as well. It would relieve the pressure at the delivery desk during busy hours, and save grown people a great deal of annoyance (Carnegie Library of Pittsburgh 1897).

As this passage illustrates, dedicated space for children was growing in perceived importance for two reasons: both to provide more effective youth services, and, also quite importantly, to diminish the extent to which adult patrons would be annoyed by young users. Indeed, Van Slyck suggests that in some cases, the latter goal dominated the former. For example, she explains how in Detroit,

the children's room was not fully integrated into the library proper. Instead, it had an outside door and entrance hall on the north side of the building, completely separate from the Woodward Avenue entrance. The purpose, as librarian Strohm reminded Gilbert in 1913, was to give the children 'no chance of over-flowing into the other part of the building.' Without the separate entrance, Strohm noted, 'I am sure it would be difficult to keep them away from other departments.' The idea was not to protect children from other library users, but to ensure that the genteel library user enjoyed the illusion of ordered and serene opulence that the architect, librarian, and library commissioners had worked to create in the stage set that they thought of as the real library (1995, 100).

Still, in other places, the inclusion of children seems to have been more mission-driven. The first children's librarian at Pittsburgh, Frances Jenkins Olcott, for example, was given her own section in the library's annual report, and used it to expound at length upon the advances in children's library services being made there, and the reasons for them. And those reasons had a great deal to do with social reform. Where earlier in the nineteenth century, libraries like Boston Public had striven to uplift public morals and attitudes through continuing education for adults, as the turn of the century approached, libraries were pursuing more of a "get 'em while they're young" strategy. As Olcott describes,

If the children of the present generation are led and taught in the right direction, the coming generation of men and women will unconsciously tend in that direction, and the evolution of reform will be natural and unforced. What greater influence is there over the mind and character of a child than the ideal he strives to follow? Nothing creates ideals sooner than books. How tremendous then is the responsibility which the public library assumes, when it attempts to guide the reading of the thousands of children in a large city (Carnegie Library of Pittsburgh 1899).

Although the reasons for the growing presumption that children would be a core library user group were diverse, the presumption itself was uniform. Children *would* use libraries, and space and resources thus needed to be dedicated to serving them (whether to keep them out of adults' hair or to serve higher educational and moral ideals). And indeed, though Carnegie himself never seems to have gotten the message about children's relevance as library users,⁸⁶ Bertram certainly did. The 1911 version of his "Notes on

⁸⁶ As of 1909, Carnegie was still describing libraries as adult-education institutions in very similar terms to those used in Boston half a century earlier:

Library Bilding" explicitly includes "comfortable accommodation for reading [books] by adults and children" among the functions to which the building should be "devoted exclusively," and in the 1916 version of the document, he further specifies that this children's reading space is among the privileged features that belong on the main floor, rather than the basement (Bertram 1911, 1, Bertram 1916, 1). And among the six plans that accompanied this document in 1916, four include explicitly sectioned-off children's spaces – and the two that do not are the very smallest, in which there is insufficient space for such divisions (Bertram 1916, 3-4).⁸⁷ As with the expression of class divides in library architecture and policies, the inclusion of children was not unique to Carnegie-donated libraries in this period; however, particularly through the institutionalized inclusion of children's spaces in the Corporation's official recommendations to communities, the Carnegie library program undoubtedly helped to make such spaces more widespread and more taken-for-granted as a necessary part of any public library.

3.1.1.3 *Racial Divides*

Race likely represents the issue of starkest dissonance between Carnegie's own views of library users and those of potential recipient communities, particularly with regard to African-American populations in the South. Carnegie himself was strongly and vocally in favor of education for African-Americans, including self-education in libraries (e.g., "Andrew Carnegie on the Negro's Progress" 1910, Nasaw 2006, 714-15). He also had absolutely no problem making donations for integrated libraries or for dedicated library spaces for African-American users (the early twentieth century being a peak time for segregation in America); indeed, he welcomed the opportunity to do so. For example, when Atlanta requested a "Negro branch" for their library system in 1904, Bertram replied on Carnegie's behalf:

Mr. Carnegie ... asks me to say that it will give him the greatest pleasure to comply with your request for Ten Thousand Dollars to erect a Free Library building for the colored people, - the City of Atlanta to maintain same at a cost of

We give to the young in the schools the rudiments of education; teach them how to use the tools by which knowledge can be acquired; we give them a taste of the dainties that are in books then as they reach the age of youth, but have not attained maturity when, more than ever they need guidance, when they need access to books [& to every good thing] and should be watched, send them adrift from the public school into the world, subject to all its temptations, and deprive them of the very aids which would best safeguard them against their dangers. It is at this very point [when school influences are gone] that the Free Library appears upon the scene with healing in its protecting wings (Carnegie 1909, brackets enclose Carnegie's own handwritten edits).

⁸⁷ It should be noted that Bertram did not arrive at his "Notes" or the accompanying floor plans independently; rather, they were informed by Bertram's consultations with trusted librarians – especially Cleveland's William Howard Brett – as well as the writings of library reformers like John Cotton Dana and William Frederick Poole. Still, as Van Slyck writes, Bertram internalized this information quite deeply: "The planning principles espoused by the library profession became Bertram's catechism, and the spread of what he called 'effectiv library accommodation,' his holy mission" (Van Slyck 1995, 34-35). And part of that "effectiv accommodation" was to provide space for children.

One Thousand Dollars a year and to provide a site for the building. Mr. Carnegie hails this as a good omen for the progress of the South, in the welfare of which he is so deeply concerned ("Free Public Library: Atlanta, Georgia" 1898, letter to J. R. Nutting, November 15, 1904).

Still, Carnegie's concern for the welfare of the South did not extend as far as using his library program to advocate for greater integration there. Neither Carnegie nor Bertram ever insisted on desegregation or equal facilities for blacks (Bobinski 1969, 80, Jones 1997, 36) – although doing so would actually have been in keeping with the *de jure*, if not *de facto*, federal law of the time (Gleason 1941, 45-46). Instead, as Jones notes,

When a town applied for a grant for a library that would be racially segregated, Carnegie and Bertram tried to compute grant amounts according to the number of people permitted to use them. Needless to say, this was not easy to estimate or enforce (Jones 1997, 32).

Population counts reported to Carnegie and Bertram by the communities themselves were unreliable in general, and this issue was compounded for estimations of the percentages of the population that were white vs. African-American. Similar issues with population estimates arose when attempting to calculate grant amounts for dedicated Negro branches, as in the case of Atlanta ("Free Public Library: Atlanta, Georgia" 1898, correspondence between Librarian Tommie Dora Barker and Bertram, June-September 1920).

The Atlanta case also provides a clear illustration of the internal community conflicts with regard to segregated vs. integrated libraries. The earliest letter to Carnegie from Atlanta, from Walter Kelly of the Carnegie Steel Company of Atlanta, suggests that given the development of new Reconstruction-era services for blacks, it would be only fair to restrict the use of the Carnegie-donated central library to whites only:⁸⁸

It is [a] noticeable fact that all the money donated for educational purposes in Atlanta by public spirited men in the North has been entirely for the benefit of the colored race, and not a dollar for the white population; when it is remembered that for years after 1868 the white population was struggling with great burdens, a kindly interest shown in this direction would have been gladly welcomed ("Free Public Library: Atlanta, Georgia" 1898, letter dated September 17, 1898).

Shortly thereafter, a supportive letter to Kelly from Clark Howell, editor-in-chief of the *Atlanta Constitution*, goes well beyond this, to declare that integrating the library would actually destroy the institution's capacity for good:

⁸⁸ This assertion, to be clear, fills me personally with a sense of futile rage. The very idea that the relative pittance that had been spent on educational and other facilities for African-Americans by northern philanthropists during Reconstruction could in any way begin to fill the gulf in opportunity between the races in Atlanta seems ridiculous on its face. And in turn, the claim that not only had this gulf been filled, but that whites therefore deserved more philanthropy dedicated to them, out of "fairness," is truly outrageous.

To throw open the Young Mens' Library Association, or any similar institution for the general use of white and black would be to paralyze the capacity it now enjoys for doing good. It would mean that the whites would be forced to give it up. This statement may seem unreasonable if viewed from the standpoint of Pittsburg, but when regarded in the light of actual conditions existing here and everywhere else in the South, it will not be misunderstood. You have been here long enough to know that what I say is true, and that it is simply out of the question to eliminate the color line ("Free Public Library: Atlanta, Georgia" 1898, letter from Clark Howell to Walter Kelly, October 29, 1898).

Still, Howell does add a vote of support for the development of *separate* library services for Atlanta's black population – though he vastly underestimates the challenge in bringing such services to fruition. As he puts it:

I have no doubt that under certain conditions the city of Atlanta would cheerfully assist in maintaining an independent institution for the use of the colored citizens, the same to be an entirely different institution, and made available for use only by colored people. It seems to me that the matter could be very easily worked out on this basis ("Free Public Library: Atlanta, Georgia" 1898, letter from Clark Howell to Walter Kelly, October 29, 1898).

At least a few letters from black citizens of Atlanta also arrived in New York in this period, pleading with Carnegie to advocate for the integration of the original, central library, rather than supporting the construction of separate facilities for the black population ("Free Public Library: Atlanta, Georgia" 1898, letters from Carrie E. Young, February 23, and J. W. B. Bowen, March 8, 1899). These requests proved fruitless; in fact, the central branch was not integrated until many decades later, in 1959 (Hein 1972, 205). Funding for a separate "negro" branch, requested through official channels in 1904, was happily granted by Carnegie within a matter of months (as noted earlier). But even then, due to an immense degree of foot-dragging, confusion, and personnel turnover on the part of Atlanta officials, construction was not even begun on the branch until 1920 – sixteen years after the granting of the original request ("Free Public Library: Atlanta, Georgia" 1898, letter from assistant librarian [name illegible] to James Bertram, November 11, 1920).

As with the other features of the Carnegie library user base discussed here, the race issues faced by Southern Carnegie libraries were fully in line with broader trends in libraries at the time – and even for decades to come. As enumerated by Gleason, even as late as 1939, in thirteen core Southern states, "the per cent of the white population receiving service [was] twice as great as the Negro population receiving service," and the "disparity between service to urban and rural Negroes is proportionally much greater than the similar disparity between the urban and rural population throughout the nation as a whole" – though 56% of the urban African-American population had access to library services at that time, in rural areas the figure was a miniscule five percent (Gleason 1941, 38-39). The problem of insufficient library access for Southern black populations thus formed an issue considerably

beyond the Carnegie program, in both institutions and time.



Two key themes run through these visions of the user in Carnegie libraries. First, the forces restricting access to Carnegie libraries, in general, were not substantially different from the forces restricting access to any other sort of libraries in similar geographic locations. And second – and closely related – the central administration of the Carnegie program had a miniscule amount of influence over the access policies of the libraries once donated – really, virtually none at all – except in the very earliest donations in Pennsylvania. The Carnegie library program as a whole was characterized by an extremely pared-down, hands-off approach, entirely lacking in oversight beyond the point of donation – at least, until Alvin Johnson was sent around in 1915. Indeed, this hands-off approach was so deeply entrenched, especially with Bertram, that when Johnson’s report called for greater oversight, Bertram’s response was to excoriate the scholar for attempting to introduce “the waste of centralized bureaucratic control,” by proposing a plan that “would require twelve secretaries in six rooms – a big unnecessary expense” (an accusation to which Johnson responded, quite reasonably: “Not so huge an expense...for keeping track of an investment of fifty millions”) (as cited by Jones 1997, 101-02). Greater oversight may well have allowed Carnegie and his associates to wield greater influence in this area – the early donations, on which Carnegie had his eye, certainly seem to have evinced a stronger desire to please the donor by complying with his vision of effective service for the working classes – yet, even were this the case, it is not entirely clear that Carnegie had any truly specific idea of what his own intended users actually looked like, beyond being willing and able to help themselves and to work for their own advancement. If anything, I suspect Carnegie’s ideal library user base would have included fewer boundaries based on physical characteristics like race – but perhaps more boundaries based on less tangible qualities, such as worthiness and aspiration.

Still, the Carnegie program did reinforce social inequalities in other ways, tied not to individual library access policies, but to the overarching administrative processes through which the program was run. These issues will be detailed below, in section 4.2.

3.1.2 Quantity

Whatever boundaries were placed on the specific user bases of Carnegie-donated libraries, it would be difficult to deny the vastness of the program’s contribution in increasing access to libraries across the United States and beyond. In the early 1920s, William Learned calculated that more than 35 million people “may be said to have access to library service from Carnegie buildings” – and even this vast figure “excludes any participating rural population not reckoned with the civil community, and in California particularly takes no account of the fact that many of the small-town libraries are centers of organized county

library service" (Learned 1924, 72). And given that the U.S. population at the time was just under 106 million, the extensive reach of the program becomes especially clear. In fact, as of the time of Learned's writing, a greater percentage of the U.S. population (31%) had access to Carnegie-donated libraries than to libraries funded in any other way (23%) – though both figures were still dwarfed by the number who had no access at all (46%) (Figure 27) (Learned 1924, 72).

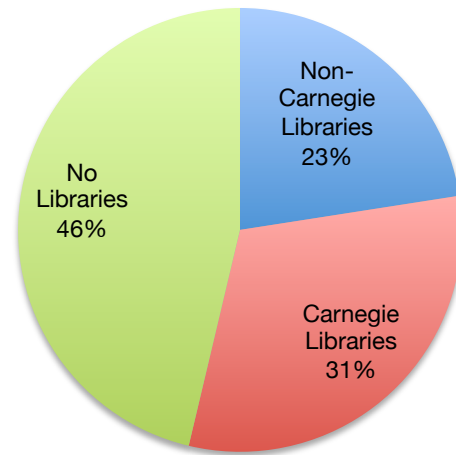


Figure 27: Access to Library Services in the United States, as a % of the Total U.S. Population, 1923

As Bobinski notes, public libraries were already proliferating at an accelerating rate across the country in the late nineteenth century: between 1848 and 1875, 188 libraries

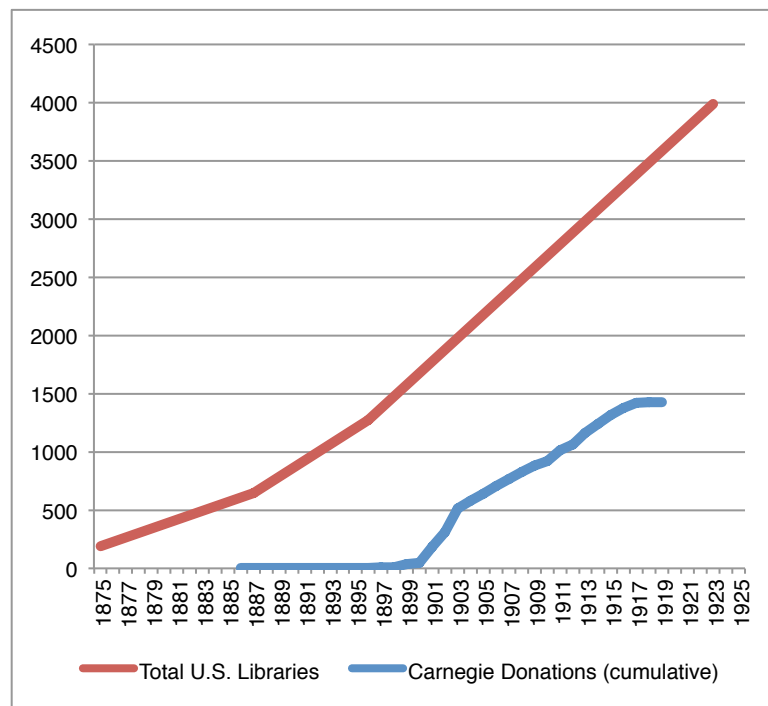


Figure 28: Approximate Growth in Number of U.S. Libraries, 1875-1923

line in Figure 28 illustrates this progression in a rough way, using Bobinski's numbers for the dates between 1875 and 1896 (1969, 7), and Learned's for 1923 (1924, 71-72).⁸⁹ It is of

were established; by 1887, the total had risen to 649; and by 1896, just before the start of Carnegie's wholesale period of library giving, there were 971 public libraries with collections of at least 1000 volumes, plus hundreds more with smaller collections (Bobinski 1969, 7). Into this upswing leapt Carnegie, providing jet-propulsion to the expansion of library services across the country, in much the same way that Joshua Bates's donations super-charged the growth of the Boston Public Library collections half a century earlier. The top

⁸⁹ The data for the cumulative number of Carnegie libraries (the lower, blue line) uses the same data as Figure 26, above. Also, for the purposes of the red line in the chart, I added 300 to Bobinski's number from 1896, to account for

course impossible to say how many of these libraries would have been constructed anyway, regardless of Carnegie's involvement. But simply looking at the number of buildings for which the program was directly responsible – and the millions of people potentially served by those buildings – the impact seems quite considerable.

3.2 *Collections*

There is a similar disconnect between Carnegie's intentions and his library donations' execution for collections as there is for users. Carnegie had many thoughts on books, literature, and the benefits of reading particular sorts of things; however, he took no role in constructing collection development strategies or policies in any libraries beyond the very earliest – and even in those, he seems to have been fairly hands-off. On the whole, Carnegie and Bertram exercised much less influence over the shapes of library collections than the shapes of the buildings that housed them. Alvin Johnson suspected that this laissez-faire attitude arose from Carnegie's belief that the specific composition of library collections didn't really matter, because providing mediocre books would help readers learn to recognize their mediocrity. As Bobinski describes, based on direct personal interactions with Johnson in the 1960s,

Johnson...believed that Carnegie did not care about the kind of books which were stocked in his libraries. Those in demand and those selected by librarians would do. An eager and persistent youth could find enough to suit his needs in any collection of thousands of titles. The philanthropist seemed unconcerned about the existence of multitudes of mediocre books on the shelves of the gift buildings. He was content to leave the fate of bad books to the law of natural selection. Carnegie once declared that "every free library in these days should contain upon its shelves all contributions bearing upon the relations of labor and capital from every point of view – socialistic, communistic, co-operative and individualistic: and librarians should encourage visitors to read them all" (1969, 109).

This loose view of library collection planning lines up squarely with Carnegie's somewhat naïve faith that public library users and librarians would be much like himself, as discussed earlier. In essence, he seems to have just assumed that once a community had a building, (a) the community would naturally fill it with books, and (b) those books would, like Colonel Anderson's collection, present a relatively well-rounded selection of information, at least partially composed of the better books of the day. In practice, however, neither of these assumptions held up well. The balance of this section will describe, first, the structurally-enforced difficulties that local communities faced in trying to build their initial library collections, and second, the divergences between Carnegie's views of particular sorts of reading and those of the library leadership in recipient communities.

some of the "hundreds" of libraries with smaller collections; 300, however, was an arbitrary number, and the reality could have been somewhat lower or quite a bit higher.

3.2.1 Size

The problem of how to pay for collections to fill the buildings was widely felt to be one of the great failings of the Carnegie program. As previously noted, the problem is well outlined in Johnson's 1916 report, where he breaks down the costs involved in running any of the smallest Carnegie libraries in detail. These libraries, built using \$10,000 from Carnegie, required their communities to pledge just \$1000 per year to support them. However, as Johnson notes, this amount was insufficient even as an ongoing maintenance amount, to cover staff salaries, book replacement and rebinding, heat, lighting, and so on – and it fell vastly short of the amount required to assemble the library's resources in the first place. In fact, this type of library, he says, had to make many unfortunate sacrifices in order to fund its initial collection:

Such a library opens its doors with few, if any, books, and it is forced to provide them out of the annual income. By limiting the number of hours when the library is open to readers, the items of heat and light can be cut, and possibly also the item of librarian's salary. But this would mean to condemn the library to a long period of stagnation, while it is accumulating a sufficient stock of books (Johnson 1916, 52).

Moreover, he continues, these sacrifices significantly imperil the library's potential for success as a community-supported institution:

This initial period of stagnation has more serious consequences than may at first appear. The erection of a library building in a small town is an event of great local importance. All the citizens take note of the progress in construction, and are in a mood of receptivity toward library propaganda. But the popular interest flags when the library remains for months or possibly years with scarcely any books, and can not even offer reading room privileges except for brief periods that are sure to be inconvenient to many. When at last the library has overcome the defect of lack of books, only by determined effort can it recover the popular standing it might have had at the outset (1916, 52).⁹⁰

Because of these issues, Johnson recommended that all further Carnegie library donations ought to take into consideration the community's ability to pull together an initial collection before making a grant, and rework the funding formula case-by-case, if necessary:

Any community manifesting an unwillingness to raise, by private subscription or public grant, sufficient funds to provide a reasonably satisfactory initial stock of books, independently of the stipulated maintenance fund, might properly be regarded as scarcely worthy of a donation. If on investigation it appeared that any community, while deserving, was too poor to make such provision for books, the question might well be raised whether it would not be wise to distribute the donation between building and books (1916, 56).

Of course, this recommendation met the same basic fate as the rest of Johnson's report: the

⁹⁰ Or, more concisely: you don't get a second chance to make a first impression.

Carnegie Corporation ceased the program rather than putting in the time and effort necessary to overhaul its procedures.

Still, Johnson's recommendation raises two interesting points, despite not having been implemented. First, it echoes the rhetoric of personal worthiness and desert that resounds throughout the Carnegie program, and indeed, throughout much of the American Public Library Movement. If communities were simply *unwilling* to provide funding, they were not among Carnegie's "deserving" class, of those willing to help themselves; however, by contrast, a willing but *unable* poorer community might make the grade, because despite lack of resources, they were willing to strive. And second, it highlights the extent to which the Carnegie library program was designed and implemented prior to its leadership – that is, Carnegie and Bertram – learning anything of substance about library administration.⁹¹ The idea that a tiny, relatively poor community might not be able to assemble a collection as broad or as useful as Colonel Anderson's, while still paying all of the library's myriad other bills in the early years, seems never to have occurred to either Carnegie or Bertram. Yet, in fact, *because* of the lack of funds for initial collection development, these small libraries often resorted to tactics like charging a one-book admission fee to a library-benefit party or social, or simply going door to door asking for books people were willing to part with (Jones 1997, 42) – strategies which tended to produce idiosyncratic collections with glaring content gaps, some of which, like an absence of practical materials, severely limited the utility of the institution (Macleod 1968, 77-80).

3.2.2 Composition

Given the difficulties in collection development faced by the recipients of Carnegie's later library donations, his statements to the recipients of his earlier, much more comprehensive and generous, donations come to seem somewhat poignant. After all, the main reason these libraries could contain the vast intellectual resources whose virtues Carnegie extolled was that Carnegie himself had agreed to personally endow them (e.g., Bobinski 1969, 82, Jones 1997, 9). For example, in his dedication of the library at Braddock, Pennsylvania, Carnegie declared,

Now I say here to the man who is ambitious to learn, who, perhaps, thinks that he has some improvement in his mind, here in this room there is, or I hope soon will be, the whole world's experience upon that subject brought right before you down to a recent date. In any question of mechanics or any question of chemistry, any question of furnace practice, you will find the records of the world at your disposal here. If you are on the wrong track, these books will tell you; if you are on the right track, they will tell you; if you are on the right track, they will afford you encouragement (1889b, 24-25).

⁹¹ This point is echoed by Macleod, in his assertion that the 10% requirement had no basis in any actual consideration of libraries' budgetary needs (1968, 50).

Ideally, Carnegie would have liked to see all of his library donations contain a diverse mix of resources: practical and instructional materials, recent news, varying perspectives on politics and economics, and, interestingly, a substantial amount of fiction (Carnegie 1889b, Carnegie 1891, 32-34, Carnegie 1894, Carnegie 1898a). However, these views were only directly communicated to recipients of the early donations, in Pennsylvania and Scotland, where Carnegie appeared personally to deliver dedication speeches, and also in some cases involved himself directly in library administration ("Correspondence" 1890-1916, "First Annual Report, Allegheny" 1891, Carnegie Library of Pittsburgh 1897). Later recipient communities likely had little idea what Carnegie himself thought they should be collecting, and even less reason to care. After all, during the wholesale period, it was made abundantly clear to recipient communities that they would receive no funds for books,⁹² and the only contact communities were generally allowed to have with Carnegie or Bertram was through the mail. Direct, in-person oversight was absent throughout, and after the money was disbursed, even indirect oversight disappeared. As such, what reason would recipient communities have had to even begin to consider Carnegie's thoughts on collections? Especially given the inadequacy of initial funds for collection building as just discussed, it seems clear that the answer is, "no reason at all." Instead, collection building among the smaller, later libraries was catch-as-catch-can – what Macleod calls the "gleanings of local attics" – often with a dearth of practical information, an overrepresentation of fiction, and no funds available to improve the balance (Macleod 1968, 77-80).

One aspect of Carnegie's views on collections, however, deserves a brief mention, despite the limitations on his influence in this area, as it serves to differentiate him quite sharply with most other early library leaders, including those in Boston. That is, his views on fiction. Where in Boston, fiction was seen as a necessary evil – something to lure the lower classes into the library, where they could be guided toward better, more substantial reading – Carnegie strongly believed in the merits of fiction in its own right, and in the absolute propriety of including it in any library collection. As he told a crowd in Inverness, Scotland in 1890:

It is sometimes urged against [public libraries] that the people who patronise them mostly read works of fiction. I am always glad when I see statements which confirm this. To the average poor man and woman, harnessed to daily toil from morn till night, I think nothing in literature is more valuable than the best class of fiction...for it lifts them out of the prosaic, monotonous life into the higher regions of imagination, and quickens their finer sensibilities....If they are denied access in the body to material palaces, I rejoice to think that through the aid of the highest works of fiction they can enter and revel in the mind through palaces and gardens and fountains, painted by master hands far excelling in beauty those

⁹² Not that this stopped many communities from trying to request additional funds. However, such requests were rarely successful in these later years (Bobinski 1969, 66-69).

that exist in reality. The working man or woman, or the young boy or girl, sitting in the flesh at the hearth of the lowly, but, surrounded by poverty, yet in the mind, the real self, thus becomes an inmate of palaces grander than those of Aladdin (1890).

For Carnegie, who had a strong love for Shakespeare and Burns instilled in him from early childhood, fiction was an escape for the mind; an opportunity to look beyond the limitations of one's everyday existence, and perhaps dream something bigger; it was a gateway to aspiration.⁹³ And although his beliefs on this issue may not have had – and in fact did not have – much impact on the perception of fiction's value within the libraries he donated, the fact he held such views at all remains an intriguing historical fact.⁹⁴

4. Implementation

While Carnegie and Bertram (and later the CCNY) may have had little direct influence over the types of users and collections that were welcomed into the libraries they donated, they had near-total control over the procedures through which those libraries were requested and granted, and, as an element of those procedures, exercised a gradually increasing degree of influence over the libraries' physical structures. Their sway in these areas, moreover, gave them fairly significant *indirect* influence over the shape of the libraries' user base and collections. The procedures and policies through which grants were made, to the extent that they empowered only specific segments of the population to engage, had an effect on who could most easily use the library and what they would find there. And the structural features and placement of buildings sent messages about what kind of institutions they housed, and who might be welcome there, even as they sketched the physical outlines for what was possible in terms of collection contents and size.

This section will discuss the role of central control in some of the processes and structures through which the Carnegie library program as a whole was implemented, and will highlight a few of the ways in which those structures and processes cycled back to influence, reinforce, or complicate the definitions and boundaries discussed in the previous section. The ordering of these sections will reverse that of the previous chapter: since the alterations in structural recommendations derive in large part from the changes in donation request processes, it will be clearer to discuss processes first, and then structures.

⁹³ As an editorial aside, I kind of love Carnegie for these views, because I agree with him. I think pop culture and pure entertainment tend to get a bad rap, even though they are basically a human universal. Fiction – be it book-based or otherwise – allows us to inhabit other lives, and thereby to escape our own for a bit. It extends our personal realities. And that stays true even if other people think it's crap.

⁹⁴ And further, as discussed in section 2.2.2, above, Carnegie's strong love for books, emphatically including fiction, was a core motivator behind his library philanthropy. So even if that love didn't seep down to his recipients, it affected their lives in a substantial indirect way, by inspiring him to give them library buildings in the first place.

4.1 Processes

The processes through which library donations were granted in the Carnegie program had wide-reaching implications for the ways in which those donations played out on the ground. The central policies put in place by Carnegie and Bertram affected whether sufficient funding would actually be made available in a given location, which groups could have their voices heard in planning and requesting libraries, and eventually, even the physical form of buildings deemed suitable for approval. These policies and procedures, moreover, evolved a great deal as the program proceeded. From 1881-1897, when Carnegie himself was more directly involved, the donations had a quite personal, paternalistic character and fairly loose requirements. After about 1898, however, with the program under the primary custodianship of James Bertram, nearly all personal sentiment was excised, and very specific and consistent conditions began to accompany donations. This section will lay out, first, some of the early experiences with library giving that helped motivate this later shift in approach, and second, the particular processes through which Carnegie and Bertram corporatized their model of library giving.

4.1.1 Early Paternalism

In the early years of the Carnegie library program, as previously noted, Carnegie tended to play the role of generous Uncle Andrew, tossing out library-shaped goodies to places that he cared about for various reasons. And given that mentality of philanthropic indulgence, the idea of limiting the scope and cost of those buildings does not appear to have been a guiding concern. Indeed, most were quite extravagant and costly. As Jones notes,

Allegheny...received an impressive five-story Romanesque-style building at a cost of \$481,012 that had an art gallery and an adjoining twelve-hundred-seat music hall with a \$20,000 organ. The Braddock building cost \$357,782 and had a swimming pool, public baths, bowling alleys, an art gallery, a gymnasium, and a billiard hall....Both buildings were many times more costly than similarly sized communities would receive ten years later (1997, 9).

And further, these libraries often also received endowments from Carnegie for ongoing support (Bobinski 1969, 76-77, Jones 1997, 9). They were essentially palaces bestowed from on high, requiring very little independent investment on the part of the local community – or at least very little explicit commitment thereof.⁹⁵

Yet, these early extravagances, so vividly contrary to Carnegie's belief that philanthropy should give "nothing for nothing," would ultimately serve to confirm the wisdom of that philosophy, with significant implications for later donations. As Bobinski notes,

⁹⁵ Library postcard images in Figure 29 and Figure 30 were retrieved from <http://librarypostcards.blogspot.com>. Cost figures in captions from Jones 1997.

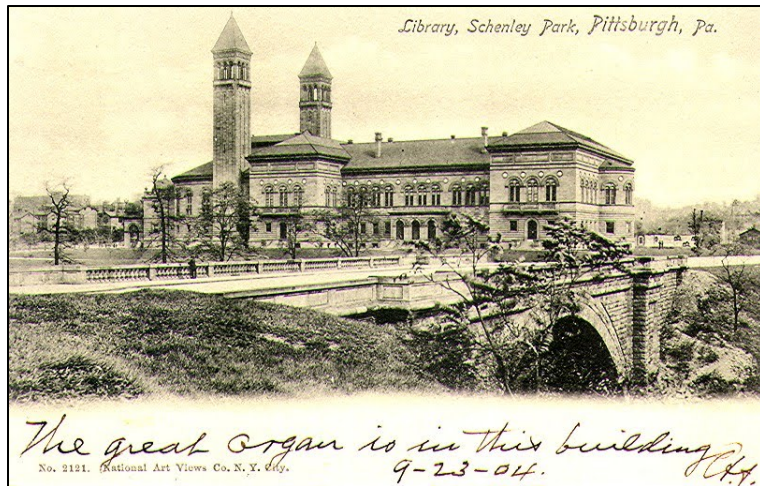


Figure 29: The Pittsburgh Carnegie Library (\$1,160,614)

rule, and another against building libraries in combination with community centers, when he gave library funds to special Pennsylvania towns which were of a personal or business interest to him and his iron industry. He lived to see these broken rules create the very situations he predicted (1969, 76).

As the years went by, these early recipient communities proved especially prone to returning to the donor with palms open, having failed to secure sufficient local funding to maintain their donated library (e.g., "Correspondence" 1890-1916, "Free Public Library: Carnegie, Pennsylvania" 1898, Bobinski 1969, 76, Black, et al. 2009, 116). Indeed, Bertram would later describe how in those years, "almost every community which received a donation from Mr. Carnegie in years gone by to erect a library building, came back with the plea that they had used the money in the building and had no money left to purchase bookstacks and furniture" (Bertram, letter to L.K. Johnson, February 8, 1916, as cited in Van Slyck 1995, 34). Such overruns and shortfalls, occurring in temporal tandem with labor-led controversies over the propriety of accepting Carnegie's "blood money" for libraries, led Carnegie to take a critical look at the conduct of his library philanthropy, and helped spur his turn to the more dispassionate system of donations that characterized the wholesale period (Van Slyck 1995, 21-23).

4.1.2 Corporatization

Though it is relatively seldom discussed from this perspective, the alterations in policy made over the

Andrew Carnegie once wrote that an endowed institution was liable to become prey of a clique, and the public might never acquire a sincere interest in it. He compared an endowed library to an endowed church, at best half and generally wholly asleep. People would never appreciate what was wholly given to them as much as that to which they themselves contributed. Yet Carnegie broke this

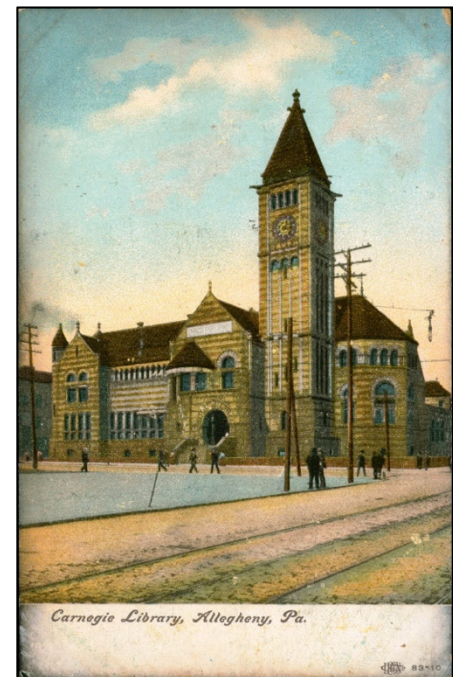


Figure 30: The Allegheny Carnegie Library (\$481,012)

course of the Carnegie library program ushered in a host of reforms to the process of philanthropic giving in general. Among those who do highlight these contributions is Abigail Van Slyck, who describes the shift in library-giving strategy as a switch between metaphors, from family to corporation. In this period, she writes, Carnegie

[jettisoned] the family model that had supported his earliest benefactions, [and] embraced the corporation as the driving metaphor for the entire philanthropic enterprise. Applying the principles of efficiency that he had developed for his railroad and manufacturing concerns, Carnegie centralized decision-making, regularized procedures, and limited the possibilities for making mistakes. Instead of becoming personally involved with the administration of his philanthropies, Carnegie established procedures that allowed others to carry out his policies. Abstract, quantitative criteria (which could be applied by anyone) replaced subjective judgments (which could be made only by the philanthropist himself) (1995, 23).

Indeed, these reforms were the beginning of what would later become legally established as the Carnegie Corporation – the first modern charitable *foundation*, though it pre-dated the use of that term (Van Slyck 1995, 24). Two changes in practice introduced during this period merit further discussion here, as they set particular boundaries on what could be requested, and by whom, within the library program: first, the systematization and bureaucratization of the procedure for requesting a library, and second, the circumscription of the program’s vision of “public demand” – a fundamentally important aspect of Carnegie’s library giving, yet one whose expression was increasingly limited to particular types of individuals over the course of the program.

4.1.2.1 Formalization: Conditions and Paperwork

After 1897, the guiding idea behind the program, that local communities would need to commit to supporting the library they were receiving, became much more formally reflected in policies and procedures, and the conditions under which libraries were granted thus became much more consistent. Early in this period, it was decided that rather than relying on informal agreements, each community would have to make a written commitment to provide a site for the library building as well as annual tax support in the amount of at least 10% of the total amount to be donated (Koch 1917, 11-12, Bobinski 1969, 187-88). And once these clear-cut, systematic requirements were in place, the path was clear for Carnegie to more fully delegate the workings of the program to Bertram, who would go on to more fully systematize and scale up the program from that point forward (Hendrick 1927, 6-8, Jones 1997, 28-9).⁹⁶ In this delegation of the detail work, the library program

⁹⁶ This transfer of power is observable within the microfilmed library correspondence: there is a marked drop-off in letters personally signed by Carnegie to any recipient community at about the turn of the century; after that point, nearly all letters responding to library requests and other related questions come from Bertram. Although Bertram rarely hand-signed any of the letters kept on file at the CCNY, his contributions are readily identifiable. Many have

became much more like Carnegie's other business enterprises in structure. As Carnegie's biographer recounts in his notes from his interviews with Bertram, Carnegie

was not a detail man; his method was always to get the right man to do the work, give him free scope and hold him responsible; he regarded his success, in business and in all other activities, as due entirely to his skill in picking the inevitable man. Thus he left this work entirely in Bertram's hands (Bertram says this himself, and I know from plenty of other sources that the statement is true) (Hendrick 1927, 7, parenthetical in original).

Still, the formalization of the site provision and 10% tax requirements were far from the end of the systematization of the program. As the donations progressed, many additional strictures were gradually added as Bertram began to encounter various problems, from towns exaggerating their population numbers to architects designing library buildings that either could not be fully constructed with the money allotted or wasted large amounts of space on things like grand staircases and lengthy hallways – or both (Bobinski 1969, 35, Van Slyck 1995, 34-35).

Many of the procedural changes Bertram introduced were geared toward making the program more data-driven: relying on official census figures rather than rough population estimates to calculate donation amounts; forcing communities to specify how they had already been spending library funds, if they had them; and so on. These shifts become especially clear through examination of the succession of forms Bertram provided to communities wishing to request Carnegie library funds (Figure 31 & Figure 32). Early on, as in the request for Ballard, WA from 1902 (Figure 31, left), the information requested was quite minimal. It asks for:

- The name of the town,
- Population,
- Whether the town already had a library (and if so, how much rent was paid for its space, and how much tax money was allocated to its support)
- Amount the community would be willing to guarantee in tax support for a Carnegie-donated building,
- Whether a site was already available, and
- How much money the community had already been able to raise ("Free Public Library: Ballard, Washington" 1903, Bobinski 1969, 203-04).

"P. Secretary" ("personal secretary") typed at the bottom in place of a signature, but even those that do not are distinguished by their terse style of writing, simplified spelling (after about 1906), and the apparently non-white paper on which they were typed, carbon-copied, or mimeographed (which comes across as various shades of gray on the microfilm).

FREE PUBLIC LIBRARY.

Town? *Ballard, Washington*

Population? *10,000*

Has it a Library at present? *Small*

How Housed? *In rooms of Business Men's Association*

Amount of Rent Paid, if any?

Amount Taxes Paid by Community, Yearly for *nothing*

Support? *By Business Men & W.C.T.U. \$30 per month*

Amount Guaranteed from Taxes Yearly if Building Obtained?

Two Thousand Dollars

Is Requisite Site Available? *Yes*

Amount now Collected toward Building? *nothing*

To facilitate Mr. Carnegie's consideration of your appeal, will you oblige by filling in the above, and returning with statement of any other particulars likely to assist in making decision?

Respectfully,
Jas. Bertram,
Secretary.

CARNEGIE CORPORATION FREE PUBLIC LIBRARY

1. Town? *Olympia, Washington*

2. Population? *Population estimate 10,000*

3. Has it a Library at present? *Yes*

4. Number of Books (including Government Reports)? *1,308*

5. Circulation for past year? *875 volumes per month*

6. How is Library housed? *Room in City Hall*

7. Number and measurements of Rooms, and their uses? *One room 36 ft long, 16 ft wide.*

8. Finances according to last Yearly Report.

Receipts:		Expenditures:	
City Appropriation,	\$ 26.00	Rent,	\$ None
Donations,	None	Salaries,	96.00
Miscellaneous Receipts,	36.00	Books,	93.65
		Miscellaneous,	106.35
Total Revenue for year,	\$ 296.75	Total Expenditures,	\$ 296.00

9. Rate and Amount which Councils will pledge for support of Library yearly (levying tax for purpose) if Building obtained? *Annual \$1500. County Commission has budgeted \$1500. Total \$3,000.00.*

10. Is requisite Site available? *Yes*

11. Amount now Collected toward Building?

To facilitate Mr. Carnegie's consideration of your appeal, will you oblige by filling in the above, and returning with statement of any other particulars likely to assist in making decision.

This blank is sent for the purpose of summarizing information and does not imply favorable consideration of the appeal.

It is necessary to give explicit answer to each question, as in the absence of such, there is no basis for action, and the matter will be delayed pending further communication.

Respectfully,
Signature *D.A. Thompson* JAS. BERTRAM, Secretary,
Designation *City and County Librarian* 2 East 92nd Street,
New York.

Figure 31: Two questionnaires for towns requesting Carnegie Libraries, from Ballard, WA and Olympia, WA, submitted ca. 1902 and 1910, respectively (CCNY Correspondence, Microfilm Reels 2 & 23)

This set of information, however, was soon deemed insufficient, and by about 1907, the forms looked more like that for Olympia, WA (Figure 31, right), which asks much more detailed questions about the operation and finances of any existing library, uses more specific wording to ask about tax support for a new library ("Rate and Amount which Councils will pledge for support of library yearly (levying tax for purpose) if Building obtained") and concludes with this more forceful statement of the importance of the community's response to the questionnaire:

To facilitate Mr. Carnegie's consideration of your appeal, will you oblige by filling in the above, and returning with statement of any other particulars likely to assist in making decision.

This blank is sent for the purpose of summarizing information and does not imply favorable consideration of the appeal.

It is necessary to give explicit answer to each question, as in the absence of such, there is no basis for action, and the matter will be delayed pending further communication ("Free Public Library: Olympia, Washington" 1903).

By the time the third version of the questionnaire was prepared (Figure 32), it had spread onto two pages, specified that population figures (both “White” and “Colored”) must come directly from the 1910 Federal Census, and required estimates of how much income the designated tax levy would have produced over the last five years. Also, this form could not be signed by just anyone. As the instructional text near the bottom reads,

To facilitate consideration of your application, kindly fill in the above, and return with signature of local authority (such as Mayor or City Clerk, if of a city, or President or Town Clerk, if of a town), whose signature will render answer to Question 9 [on tax support] effective (Carnegie Corporation of New York).

Some of the implications of this restriction in signatory power will be discussed in the next section (4.1.2.1).

**Carnegie Corporation
of New York**
570 Fifth Avenue
New York

Free Public Library Bidding

Precise answer to each question is necessary. Please use typewriter.

- Name of City, Town, Village, Township or County which library bidding would serve _____ State _____
- Population at 1910 Federal Census: White _____
Colored _____
Total _____
(Give official figures, not estimates)
- Has it a Library at present? _____
- Number of Books (excluding Government Reports) _____
- Circulation for past year _____
- How is Library housed? _____
- Number and measurements of Rooms, and their uses _____
- Finances according to last Yearly Report.

Receipts:	Expenditures:
Tax Appropriation, \$ _____	Rent, \$ _____
Bequests, _____	Salaries, _____
Miscellaneous Receipts, _____	Books, _____
_____	Miscellaneous, _____
Total Revenue for year, \$ _____	Total Expenditures, \$ _____
- Amount which Councils will pledge from taxation for annual support of Library if Bidding is obtained \$ _____
(Give specific amount, not rate of levy)
- 9-a. (over)
- Is requisite Site available? _____
- Amount now available toward Bidding _____

To facilitate consideration of your application, kindly fill in the above, and return with signature of local authority (such as Mayor or City Clerk, if of a city, or President or Town Clerk, if of a town), whose signature will render answer to Question 9 effective.

This schedule is sent for the purpose of summarizing information, and does not imply favorable consideration of the application.

Signature _____

Official Designation _____

Date _____

JAL. BERTRAM, Secretary,
Carnegie Corporation of New York,
570 Fifth Avenue,
New York, N. Y.

- 9-a. What is the highest rate of levy allowed by law for support of library? _____
(State answer, if possible, in mills in the dollar)
- 9-b. What income would this rate of levy have yielded during each of the last five years?

Figure 32: A third version of the questionnaire, used after the founding of the Carnegie Corporation in 1911 (Carnegie Mellon Libraries, Digital Collections)

Although these procedural changes may seem minor taken individually, they had a major impact on the practical experience of requesting a Carnegie library when viewed as a

whole. As Jones notes,

depending on the date of application, two neighboring communities of similar size could have completely different experiences in applying for and receiving Carnegie library grants. In 1898, for example, a community might have communicated with both Carnegie and Bertram, quickly and easily received grant approval, haggled for a little extra cash, and been given complete freedom to construct the building it wanted. By 1914, a community had to negotiate only with Bertram, proving the need for a public library grant with substantiating data, and if successful received an amount that was rarely negotiable. Thereafter, they had to submit official community resolutions regarding the maintenance fees and building site, as well as their blueprints, all of which were rigorously critiqued and often revised (1997, 28).

All of these added formalities, moreover, were required to take place almost exclusively via mailed correspondence, per new program policy. Whereas in some of the early cases Carnegie had been willing to meet personally with local leadership and even to make occasional site visits (e.g., "Correspondence" 1890-1916), such face time was rarely granted after about 1897 (Bobinski 1969, 35; also evident from the correspondence reels, which include many letters declining to grant such meetings and visits). This placed an additional layer of detachment between the central leadership of the program and communities requesting libraries, with implications for both the objectivity and timeline of library grants (which became greater and slower, respectively).

4.1.2.2 *Authorization: Public Support, and Who Can Declare It*

Public demand was much more important as a direct enabling factor in the Carnegie library program than it had been in Boston; however, it comes up in a somewhat unusual way. That is, Carnegie structured the program in the way he did – as a “bribe” – precisely in order to encourage public buy-in and support.⁹⁷ As he put it at the dedication of the library in Washington, D.C., “This is not charity, this is not philanthropy; it is the people themselves helping themselves by taxing themselves. They owe no man anything of moment” (“Dedication of Washington (DC)” 1903). And later, James Bertram would explain to Carnegie’s biographer that his employer regarded

[the] creation of comprehensive public libraries...as the most pressing of educational needs, yet he realized that the movement must be a public one: that private philanthropy could not supply so great a public need, and that real social statesmanship consisted in organizing the existing public demand and putting vitality into it. So he hit upon his scheme to get action out of the public authorities - which in the last resort meant the politicians. This was to offer communities a building, on condition that they would fill it with books and tax

⁹⁷ In this way, it actually starts to look like early example of what we might now call “Incentive-Centered Design:” structuring the social and economic framing of a situation in such a way that individual incentives align with desirable social outcomes, and thus make them more likely to occur (e.g. <http://stiet.si.umich.edu/icd>).

themselves for its maintenance. Carnegie himself described this as a "bribe" to cities and towns to establish libraries (Hendrick 1927, 5-6, paraphrasing Bertram).

Whereas in Boston, the public authorities and particular members of the local elite were at the very forefront of the push for a public library, with the major bankroll only following later, Carnegie offered an initial bankroll – if for a building only – in order to motivate local authorities and social groups to take action. Whether or not this was a successful ploy is a matter of historical debate. Bertram decisively claimed that it was, declaring (via Hendrick) that Carnegie's "plan worked, - and the great library extension of the present time is the result. Now the library is as much an accepted part of every community as the schoolhouse - which was exactly the condition A.C. hoped to bring about" (Hendrick 1927). Later writers like Macleod, however, have suggested that the situation was more complicated: that preexisting public agitation would have led to a boom in library construction in many regions during this period regardless of Carnegie's actions, and that even where Carnegie did donate libraries, they often failed to gain a foothold in places where this kind of authentic, previously established demand did not exist (Macleod 1968, 35, 52, 64). Some have also trodden a middle ground, describing the demand for public libraries and library philanthropy as a feedback loop in which existing demand created a path of low resistance for the donor, whose donations, in turn, fostered further demand (Arthur E. Bostwick, Chief of the Circulating Department of the New York Public Library, cited in Koch 1917, 43-44). Whichever may be the case, it seems clear in purely pragmatic terms that among the later donations, none were granted without at least a modicum of expressed public demand, specifically in the form of commitments of land and tax money.

Yet, not all public voices were equally relevant in expressing this public demand. One piece of the systematization of the donation process, as just described, was to circumscribe the set of individuals who were empowered to speak for their communities in the context of library requests. And Bertram had pragmatic, experience-based reasons for making this adjustment, as described by Jones:

Bertram learned the hard way to carry out grant negotiations only with town mayors, councils, or officials appointed to represent them. In a small northwestern community, the mayor wrote Bertram for grant information only to discover that his town had already been approved for a grant after extensive negotiations with the local newspaper editor. The editor later confessed that he had told no one of his efforts so that a building site could be procured before the sellers found out about the Carnegie grant and raised the price (1997, 29).

Similar situations were not uncommon. As such, the decision to deal only with local officials originated from a reasonable desire to avoid confusion and inefficiency. However, the policy had significant unintended consequences for the local power dynamics surrounding the library requests. In particular, this policy compounded existing

inequalities for segments of the population that were already politically marginalized, including women and African Americans. Each of these merits a brief discussion here.

4.1.2.2.1 *Women*

Women often played an extremely significant role in completing the legwork involved in making a Carnegie library request – raising funds, collecting the information for Bertram’s questionnaires, even in some cases establishing small proto-libraries to prove their community’s commitment (Jones 1997, 42-43). And once the library was built, women generally comprised a significant part of its staff; by 1920, 90% of librarians were women, and that trend was well underway by the turn of the century (Garrison 1979, 173).

However, between the phase of drumming up support and that of staffing the library, the necessary involvement of local voters (i.e. men) and/or local political leaders (also men) in approving the tax basis for the library created myriad opportunities for male co-option of women’s efforts. This was only compounded by the requirement, in later years, that those local political leaders were the only ones allowed to carry on the official correspondence with Bertram. And available evidence suggests that co-opt they did. As Van Slyck writes,

Soon after opening negotiations with the philanthropist...club women found themselves in an uncomfortable situation. Although they were free to place their case before him, Carnegie was careful to formalize every step of the negotiations with the elected officials of the local municipal government. The mayor’s signature was required on the application form, the city government had to provide a library site, and the town council had to pass a resolution to dedicate city tax dollars to a library maintenance fund. Unable to hold public office, club women discovered that a Carnegie gift could mean losing control over the library they had established. Since middle-class club women, however, were often related to the civic leaders, their disenfranchisement was never automatic and rarely absolute. Nonetheless, the Carnegie system helped institutionalize existing patterns of gender relations in cultural matters, officially crediting female initiative in library matters to men, and officially recognizing female labor only when it was sanctioned by male authority (1995, 134).

Several other authors, moreover, provide further support for these contentions. Jones recounts the story of Nampa, Idaho, where the women of the Century Club successfully raised the funds for the library, only to have their control of the project usurped by a mayor-appointed male building committee (1997, 43). And Garrison notes that although most library *staff* were female, “the evidence is overwhelming that public library *trustees* have been almost entirely male, white, well-educated, well-to-do, and usually professionals or businessmen,” and that, as the profession of librarianship became increasingly feminized, there was a trend toward “subservience of librarians to their trustees,” with roots in both the social and political realities of the time (1979, 50, text and footnote, emphasis added). As previously noted, women were more than welcome in libraries as both users and as staff; many even carried on significant and meaningful correspondence

with Bertram, especially as librarians, after their local libraries had already received Carnegie grants (see, e.g., "Free Public Library: Atlanta, Georgia" 1898, correspondence with Tommie Dora Barker). At the point of meeting official legal requirements and formalizing donation arrangements, however, the structure of the program required male involvement in the process.

4.1.2.2.2 *African-Americans*

Similarly, local African American leaders in the South – as well as average African American citizens – often advocated strongly for library service for black populations, either through the integration of whites-only libraries or through the construction of separate branches; this is especially evident in the Atlanta correspondence (e.g., "Free Public Library: Atlanta, Georgia" 1898, letters from Carrie E. Young, February 23, 1899, and J.W.B. Bowen, March 8, 1899). However, there is no evidence to suggest that these Atlanta advocates, at least, ever received any direct response from Carnegie or Bertram. In Atlanta, movement eventually did begin toward the construction of a "negro branch," but only after it had been requested through official – white – channels. Most of the correspondence regarding the details of that branch donation was conducted with the librarian of Atlanta's existing central, whites-only branch, Tommie Dora Barker (and later Mrs. Blewett Lee), in 1916 and 1920 – although the initial request for funds had been granted in 1904 ("Free Public Library: Atlanta, Georgia" 1898). Malone's depiction of the opening of the first library branch for African Americans in Louisville, Kentucky presents a similar image, where black advocacy had to be heard by local white leadership in order for action to be taken, and where that action seems to have been at least as much a product of fear of integration as actual desire to provide the service: the branch was opened by the Louisville Free Public Library in 1905 (prior to the Carnegie donation for it), she writes, "mainly in response to the activism of black school principal Albert Meyzeek and out of a desire to reserve the main library for whites only" (Malone 2000a, 182). That branch later moved into a new, separate Carnegie building in 1908. And Bobinski describes a range of cases all across the South where the question of whether or not the black population would have library services was left entirely in the hands of the local white leadership, to varying effect:

Charleston, South Carolina, was reluctant to establish a free public library...because officials feared Negroes would want to use it. Union Springs, Alabama, declared that the colored people would not be allowed to use the library building, but if they expressed a desire for a library of their own, no doubt provision would be made for them. In preliminary correspondence, officials from Rome, Georgia, indicated that Negroes would have access to the library, since it would be illegal to collect public funds and not have facilities available for both races. But after the library was built, they put in an application for a Negro branch....Jackson, Mississippi, representatives were more blunt. They said Negroes were too illiterate for library facilities; the educational level of the colored people had to be raised first (Bobinski 1969, 80-81).

Understandably, as in Atlanta, other African American leaders across the South protested the inequality of library services in their areas, sometimes writing directly to Carnegie and Bertram to request dedicated facilities to be maintained through non-municipal channels. These requests, however, were not successful. Instead, they hit a brick wall of Carnegie conditions: 'money could only be given for libraries maintained by cities and towns,' Bertram would write; or, 'we do not interfere in local matters,' the CCNY would protest (Bobinski 1969, 81, paraphrased). And so the process of obtaining Carnegie library buildings for Southern African Americans, while explicitly welcomed by both Carnegie and the CCNY, tended to get mired in a swamp of entrenched discrimination and apathy – or even outright hostility – on the part of those with the power to act under the rules of the program. Just as "separate but equal" never managed to be the latter, so did the Carnegie policy of declaring their support for equal library services while only actually consenting to negotiate with local officials – who were invariably white – effectively silence the voices of African Americans within the program. While in theory, Carnegie gave libraries as a way of helping those willing to help themselves, in this case the program's own policies walled off the path to that help for a very willing segment of the public.

4.2 Structures

The structures worth discussing within the Carnegie library program are a bit different and more diffuse than those in Boston. Where in Boston, it was possible to examine a single initial building with a single internal arrangement and a single line of progression in cataloging, there is no such well-considered, unitary artifact in the Carnegie program.⁹⁸ Instead, each recipient community had its own internal discussions as to what their library would look like and what purposes it would serve, typically without very much input or interference from Carnegie or Bertram (e.g., Bobinski 1969, 57). Indeed, neither Carnegie nor Bertram appears to have gone into the program with any strong opinions about library design. As the program proceeded, however, issues of building design and placement

⁹⁸ Indeed, despite being a core structural focus for the early leadership of the BPL, classification systems are never even mentioned by Bertram or Carnegie, though they do come up in a few of the libraries' annual reports. By this time, there were standard systems widely available for adoption – namely Dewey and Cutter – so the libraries did not have to invent their own systems from scratch as Boston had. Still, like BPL, libraries like Allegheny and Pittsburgh seem to have started with printed, book-form catalogues, and only later switched to a card system. ("First Annual Report, Allegheny" 1891, Stevenson 1892, "Third Annual Report, Allegheny" 1893, Carnegie Library of Pittsburgh 1897, Carnegie Library of Pittsburgh 1898, Carnegie Library of Pittsburgh 1899). Atlanta seems to have made a card catalogue straight out of the gate; the librarian refers to it in her annual report as early as 1901: "In addition to our official card catalogue, which will be used in the reference room of the new Library, we are making a typewritten duplicate catalogue for use in the open-shelf room. This work is being done on best grade, standard sized cards, a Smith-Premier typewriter being used" ("Second Annual Report of the Board of Trustees of the Carnegie Library of Atlanta" 1901). Finally, Braddock mentions switching from Cutter to Dewey in 1908 ("Twentieth Annual Report of Carnegie Free Library and Sixteenth Annual Report of Carnegie Club, Braddock, PA, for the Year 1908, Together with Statistics from the Founding of the Library" 1908).

became wrapped up in the larger processes of systematization described above, and especially during the wholesale period, Bertram and Carnegie began to express some quite explicit views on how best to plan libraries and library systems. This section will explore this evolution in structural thinking within the Carnegie program through two basic elements: building planning and stacks arrangement. My analysis will draw significantly upon Van Slyck's much more in-depth examination of Carnegie library architecture, but will avoid fully retreading it, as I see no need to completely reinvent what has already been masterfully done.⁹⁹ In order to provide a fuller picture of the trajectory of these changes, moreover, this section (like the last) will cast its focus a bit beyond the designated time scope for this case, incorporating relevant pieces from later in the wholesale period, along with evidence from the initial retail phase.

4.2.1 Building Planning

The gradual shift to a more corporate, systematic program of Carnegie library donations, as described above, had an enormous effect on library architecture both within the Carnegie program and beyond. Along with the systematization of donation policies came a drastic reduction in the average donation amount, and with that reduction came a more pressing need for a more economical and pragmatic vision of library architecture than had previously reigned. Indeed, the early buildings donated (as in Figure 29 and Figure 30) ultimately proved entirely atypical with regard to the Carnegie library program as a whole in both budget and structural implementation, reflecting much more of the ostentation, impracticality, and inefficient design of past periods than the simplicity and economy that would later become almost synonymous with the term "Carnegie Library" (e.g., Van Slyck 1995, 1-43). This shift in library architecture took a dialogic course: the prevalent thinking among librarians and library reformers influenced Bertram's notions of ideal library design, and Bertram then utilized his gatekeeper role on Carnegie library funding to spread that vision back out to communities requesting libraries. As Van Slyck describes, "[the] planning principles espoused by the library profession became Bertram's catechism, and the spread of what he called 'effectiv library accommodation,' his holy mission" (1995, 34-35).

Notably absent in this cycle of influence were architects – and this seems to have been completely intentional on the part of both the librarians and Bertram. Librarians had long railed against architects' tendency to privilege aesthetics over pragmatism when designing libraries; indeed, the previous chapter catalogs some complaints of exactly this sort made by Justin Winsor against the BPL's Boylston Street building (Chapter 3, Section 4.1.1; also

⁹⁹ Unlike that of any of the other cases examined here, the architecture of the Carnegie libraries has already been the object of extensive social analysis in its own right, in Van Slyck's *Free to All: Carnegie Libraries & American Culture, 1890-1920*.

Whitehill 1956, 79).¹⁰⁰ Thus, in the late 19th century, there was an increasing push by librarians to involve themselves more directly in the design of library buildings, because only they could truly understand the functional requirements those buildings would entail (e.g., quote from Mrs. Percival Sneed, cited in Van Slyck 1995, 35). And as the Carnegie program progressed, Van Slyck notes, Bertram came to internalize “not only the librarians’ dogma but also their prejudices. Long considered the natural enemy of the librarian, the architect became Bertram’s personal *bête noire*” (1995, 35). And indeed, by 1916, one can almost hear the disdain for that profession in one of Bertram’s customarily terse notes, admonishing an Atlanta librarian to “Pleas note that all communications should come from civic authorities; we do not wish to correspond with architects” (“Free Public Library: Atlanta, Georgia” 1898, letter to Tommie Dora Barker, November 9, 1916).

Still, where librarians’ objections to impractical library architecture derived mainly from the obstacles it created to the provision of efficient service to the public (e.g., Poole 1881, Dana 1897), Bertram’s animus toward the same object had a different root. That is, prior to his exposure to the world of library planning, Bertram already shared his employer’s abhorrence of monetary waste, and desire to avoid such waste in philanthropic efforts (e.g., Hill 1936, 32). In keeping with this philosophy, Bertram’s zeal for efficient and effective library building had much more to do with a desire to control costs than with any thought of improving public service (Van Slyck 1995, 34, Jones 1997, 53-54). As previously noted, the early communities receiving library donations had proven prone to running out of money too quickly, and then returning to Carnegie – via Bertram – to beg for more. And there was a sense that given the requirement of public support attached to these institutions, their architecture should embody and convey a sense of fiscal responsibility (Van Slyck 1991, 370). These experiences and sentiments, in combination with his readings of the works of reformers like Poole and Dana, as well as his correspondence with Cleveland librarian William Brett, ultimately led Bertram to the conclusion that such overruns centrally resulted from poor building planning, and could be remedied through better guidance in that area (Van Slyck 1995, 34).

And indeed, details from the reports of some of the early libraries lend credence to this

¹⁰⁰ And my own experiences would suggest that this battle has never really ceased. One of the first libraries in which I myself was employed, at Northwestern University, was built in the 1970s as an interlinked set of three cylindrical towers, which respectively housed the book stacks. Beyond making it inconvenient to move between subject collections (which required return to the second floor to move between collections housed on higher floors in different towers) the shelving in these towers was arranged like spokes on a wheel, radiating out from the center on each level. As one might imagine, this made it extremely challenging to accommodate a growing collection, as things like high-density shelving – or even the addition of more standard shelving – were made impossible by the radial configuration. The architecture of the building was a frequent topic of ridicule by librarians on staff (including myself): a totally impractical design, foisted upon them by a building committee on which librarians were reportedly not granted much of a voice.

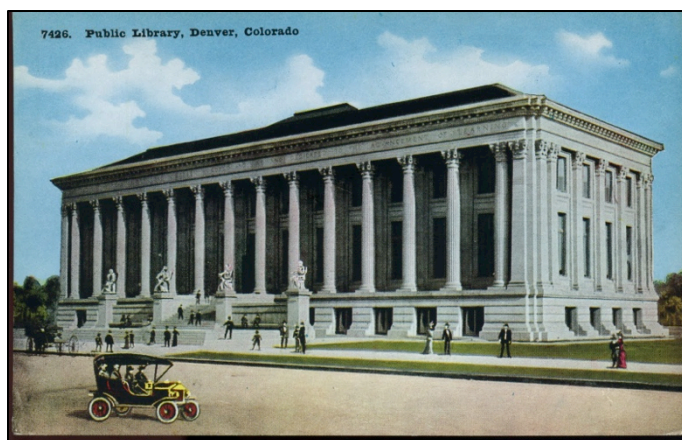


Figure 33: Denver Public Library (1902, \$200,000)

conclusion. For example, the Allegheny building was initially planned without any space for a reference room ("Third Annual Report, Allegheny" 1893, 12-13) and Pittsburgh without space for children's services (Carnegie Library of Pittsburgh 1897, 13-14) – both omissions that were immediately evident as errors upon opening those buildings. And in at least the Allegheny case, the omission was ultimately remedied through an

additional infusion of Carnegie funds ("Third Annual Report, Allegheny" 1893, 12). But these cases were only the beginning. Jones catalogs a litany of aesthetic excesses in early wholesale-period libraries, pursued to the detriment of their practical utility – from a Houston library that looked "like a wedding cake," in which "[all] the standard elements [were] present, but in multiples," to the functionless dome atop a tiny library in Greensburg, Indiana (Jones 1997, 69). At least one of these examples of inefficient excess provoked a response from Carnegie himself (though one suspects it was not alone in doing so). Upon seeing a newspaper mock-up of the new building he had just funded for Denver, Colorado (of which Figure 33 is a later depiction), Carnegie reportedly "scribbled across it: 'I am sorry to have my money wasted in this way. This is no practical library plan. Too many columns!'" (Jones 1997, 69).¹⁰¹

As of 1904, Bertram started to request building plans from communities that were running over budget, and by 1908, this sort of review had become mandatory prior to approval of any library request (Van Slyck 1995, 34). The apotheosis of this trend toward oversight and architectural review, however, came in 1911, when Bertram began to distribute his "Notes on Library Bilding"¹⁰² to prospective recipient communities. Recognizing that his existing strategy of evangelizing for efficient planning on a library-by-library basis was *itself* inefficient, Bertram condensed his accumulated views on library economy into a concise screed that grew from one page to four over the next few years. As a four-page document, two contained text, while the other two presented six sample library floor plans (Figure

¹⁰¹ Interestingly, the Denver building to which Carnegie objected was actually designed under the aegis of one of the library reformers whose writings influenced Bertram to discourage exactly such monuments – John Cotton Dana, who was then the City Librarian (<http://denverlibrary.org/content/dpl-history>).

¹⁰² Several versions of this pamphlet were produced, under slightly varying titles; I found at least two in the CCNY Archives at Columbia – from 1911 and 1916 – all jumbled together. "Notes on Library Bilding" is the title from the 1911 version; later versions seem to have gone by "Notes on the Erection of Library Bldings" (as shown on both the archival version from 1916 and the 1915 version reproduced as Appendix I in Van Slyck 1995).

34).¹⁰³ In all of its forms, the text is a paean to the relationship between efficiency in structure and efficiency in cost, while the plans, added later, provide specific, executable ideas about how this relationship might be expressed in library architecture. As Bertram declares, Carnegie library donation amounts are:

sufficient only to provide needed accommodation and there will be either a shortage of accommodation or of money if this primary purpose is not kept in view, viz.: TO OBTAIN FOR THE MONEY THE UTMOST AMOUNT OF EFFECTIV ACCOMMODATION, CONSISTENT WITH GOOD TASTE IN BILDING (Bertram 1915, all-caps section in original).

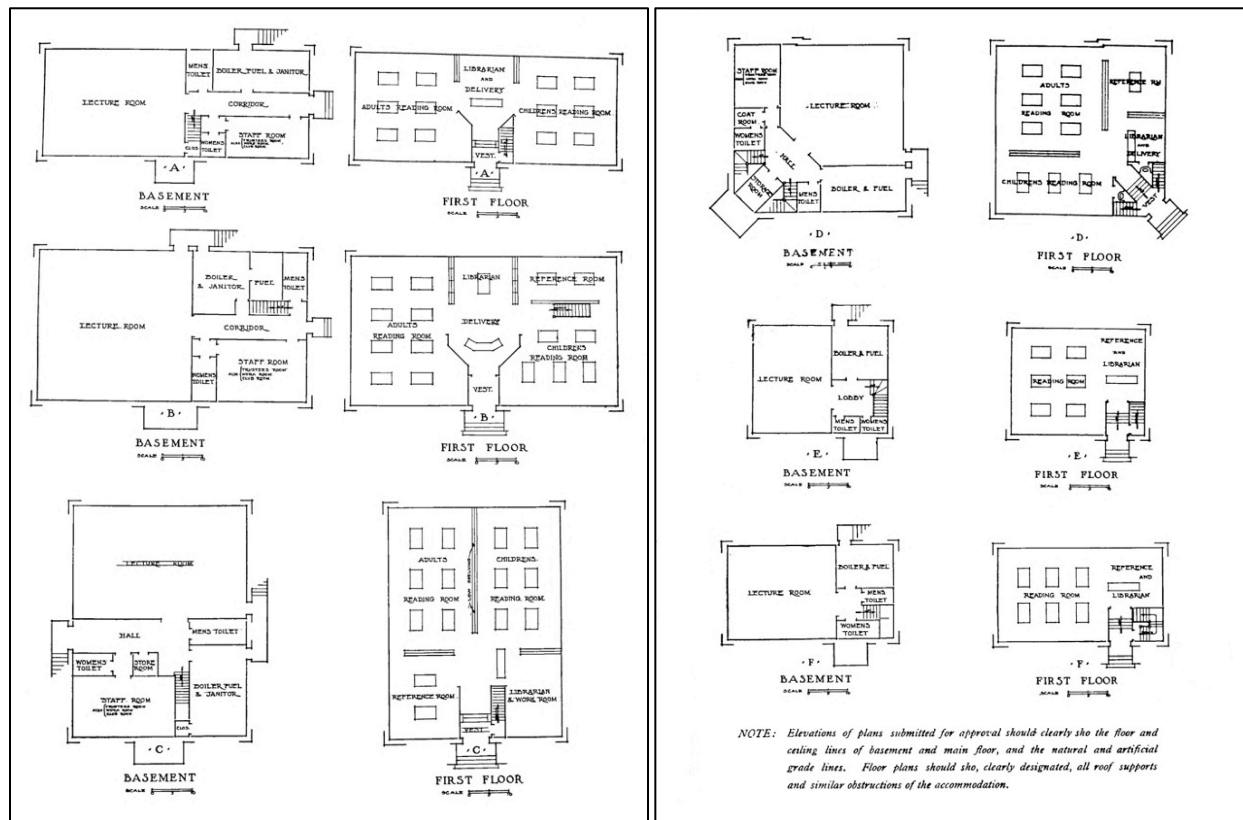


Figure 34: Sample library plans from Bertram's "Notes on the Erection of Library Bldings," 1916.

And he goes on to describe some very specific views of how this "utmost amount of effectiv accommodation" for the money might best be attained. First, he observes:

In looking over hundreds of plans for small and medium-sized bildings, costing about \$10,000, more or less, we hav noted some features leading to a wasting of space, especially in connection with the entrance feature, which, when not wisely pland, leads also to waste in halls, delivery room, etc.

The economical layout of the bilding is sacrificed or subordinated at times to

¹⁰³ As Van Slyck notes, the "Notes" were not exclusively Bertram's creation, but more likely a collaboration between Bertram, several librarians, and perhaps New York architect Edward L. Tilton (1995, 35)

minor accessories, such as too much or too valuable space allotted to cloak rooms, toilets and stairs (1915).

In fact, Bertram declares that the only purposes that a library should rightfully serve are, on the main floor, "housing of books and their issue for home use; comfortable accommodation for reading them by adults and children," and in the basement, a "lecture room; necessary accommodation for heating plant; also all conveniences for the library patrons and staff" (1915).

Given these condemnations of wasted space and restrictions on uses deemed appropriate for the building, he then goes on to provide a very specific outline for the efficient arrangement of small library buildings:

Experience seems to show that the best results for a small general library are obtained by adopting the one-story and basement rectangular type of building, with a small vestibule entering into one large room subdivided as required by means of bookcases. In cases where it is necessary, to secure quiet, glass partitions may be put above the bookcases.

By a one-story and basement building is meant a building with the basement about four feet below the natural grade, the basement being from say 9 to 10 feet and the main floor from say 12 to 15 feet high in the clear. Plans have at times been submitted for "one-story and basement" buildings, which differ from two-story buildings only by having the stair to the upper floor outside instead of inside!

The rear and side windows may be kept about six feet from the floor, to give continuous wall space for shelving. A rear wing can be added for stack-room (when future need demands it) at a minimum expense, and without seriously interfering with the library service during its construction.

The site chosen should be such as to admit light on all sides, and be large enough to allow extension, if ever such should become necessary (1915).

He also avers that "Small libraries should be planned so that one librarian can oversee the entire library from a central position" (1915) – itself a relatively novel concept in library planning, directly related to the broader adoption of open stack arrangements, which will be discussed in a moment. Absent from Bertram's dictates, however, was any mention of aesthetics (as noted by Van Slyck 1995, 36-37), aside from a gruff acknowledgement that building exteriors

are features in which the community and architect may express their individuality, keeping to a plain, dignified structure and not aiming at such exterior effects as may make impossible an effective and economical layout of the interior (1915).

The plans themselves exemplify these dictates, as is their intended purpose. All six depict libraries with only a first floor and basement level: reading rooms and service desks above, lecture halls and staff space below. All but the two very smallest plans (E & F) include

dedicated space for children, and all exhibit an open, rectangular floor plan, to maximize usable interior space. None segment women off into their own dedicated reading rooms. And although the stated purpose of these architectural reforms was to increase efficiency and reduce costs, their social consequences went well beyond those aims. As Van Slyck suggests, these plans all but flattened the architecturally-expressed social hierarchy of earlier library buildings, like those in Allegheny and Boston:

For library patrons, male and female, young and old, the new library offered a pleasant surprise. From the outside, the emphasis on symmetry helped identify the building as a public one; readers could enter freely, secure in the knowledge that they were welcome. Inside, the architectural experience had been evened out. Ceilings were of a uniform height, and rectangular rooms were lit from windows that started six feet from the floor. Gone were the specially shaped reading rooms with their aura of Victorian domesticity. Gone were monumental vistas into large public rooms. If the experience was less dramatic, it was also less intimidating (1995, 41-42).

And further, the symmetrical, open-plan library plans promoted in Bertram's "Notes" had a strong influence in library planning well beyond the Carnegie program itself; throughout the first half of the twentieth century, this was simply how public libraries would be built by default. Indeed, as late as 1941, the so-called "Carnegie Rectangle" was deemed to merit a full chapter in Wheeler and Githens' volume on library planning, which affirmed it as an example of "sound common sense" in the domain (Wheeler and Githens 1941, 215-25, Van Slyck 1991, 383).

4.2.1.1 Branch Libraries

Though perhaps not as significant a trend as Bertram's reforms of library architecture, a second element of Carnegie program building planning that bears noting was its shift away from supporting the construction of grand, central library buildings, and toward providing for the construction of networks of branch libraries. In part, this shift in thinking emerged from the same set of experiences as had the architectural reforms. As Bobinski explains, a few years into his wholesale giving period, Carnegie

changed his views on donating funds for large main libraries in big cities in favor of providing funds for branch libraries. He felt there was a tendency for such central buildings to be too lavishly designed and monumental (1969, 70-73).

Even before this, however, Carnegie had shown a strong interest in supporting networks of branches in larger cities. As early as 1890, Carnegie had written to library supporters in Pittsburgh that,

[the] experience of New York, Baltimore and other large cities has proved that a central library should be supplemented by branch libraries. The Free Circulating Library of New York has now four of these, the City of Baltimore has five; they are not extensive structures, but each contains a small supply of the books most

in demand and a Reading Room, and is operated in connection with the Central Library ("Free Public Library: Pittsburgh, Pennsylvania" 1890, letter from Andrew Carnegie, February 6, 1890).

These examples, he continued, illustrated the wisdom of establishing a set of branch libraries in Pittsburgh as well; a contention which he then proceeded to financially back to the tune of \$300,000 ("Free Public Library: Pittsburgh, Pennsylvania" 1890, letter from Andrew Carnegie, February 6, 1890, Van Slyck 1995, 102). He continued to advocate for branch service as the program progressed, in ever-stronger terms. Indeed, the largest sum of money Carnegie donated to any city - \$5.2 million for New York City in 1901 - provided exclusively for the construction of 65 branch libraries, and no larger central buildings ("Free Public Library: New York, New York" 1899, letter from Andrew Carnegie to J.S. Billings, Director, NYPL, March 18, 1901). And by 1908, Carnegie had ceased giving any money for large central libraries at all, preferring to bring "books close to the homes of the people" by donating smaller branch buildings (Van Slyck 1995, 102, citing letter from James Bertram to Harry P. Carlton of Oakland, CA, 1908). Van Slyck provides a useful analysis of the ambivalence of these branch library donations - and their donor - toward their largely working class users, as expressed through their structures: on the one hand, they were intentionally located in working-class areas for working-class use, and tended to adopt reformist open-shelf policies that allowed those users greater freedom of access to the books; but on the other, their internal arrangements often revealed distrust for those same working-class users, embodied in a floor plan that enabled constant surveillance by the librarian as well as physical control over readers' movements through the space (Van Slyck 1995, 101-10). These types of behavioral control and the classist distrust in which they are rooted, moreover, strongly echo Garrison's account of the librarian elitism more generally prevalent in the same period (Garrison 1979, 39-42). Still, in the context of the Carnegie library program, the shift toward provision of branch libraries remains basically ideologically consistent with the program's broader trend toward maximizing library economy and utility, as donating smaller branches would not only avoid much of the ostentation and waste of larger central buildings, but would also provide more efficient service, by placing the book collections in locations more proximate to their user base.

4.2.2 Stacks Arrangement

The final piece of Carnegie library structure worth mentioning, though extremely important and influential, can also be stated briefly. That is, the Carnegie library program brought with it a fairly massive increase in the prevalence of open-stack shelving arrangements across the country, starting quite early in the program. Based on a 1902 survey of library architecture, Van Slyck calculates that Carnegie libraries at that time were fully 25 percent more likely to provide open access to the book shelves than non-Carnegie libraries - although this format did not represent a majority of either type (44 percent of Carnegie

libraries used closed stacks vs. 68 percent of non-Carnegie libraries) (1995, 34). At the time, open shelving was still fairly controversial, so these numbers are quite significant.

Further, though it is beyond the scope of the current study, it would be interesting to know how that same comparison from 1902 might have looked twenty years later, given Bertram's ever-more-forceful architectural recommendations over that period. One of the striking features of Bertram's "Notes," after all, is that every one of the six model library plans presented there appears to employ an open shelving plan – indeed, it would be difficult to envision how a closed-stack area could even be made to fit in any of them.¹⁰⁴ Like the rest of Bertram's architectural suggestions, his support for open stacks likely emerged less from the high-minded faith in the capacities of the average library user expressed by reformers like Dana (e.g., 1897, 245-46), than from the very consistent ethos of economical library planning described above. Indeed, one can imagine another of Dana's claims – that the architectural model for the library should be found in the realm of business, rather than monumentalism – resonating strongly with Carnegie and Bertram, as they pioneered a more corporatized system of philanthropy and ever-greater efficiencies in library planning. The library, Dana writes,

if it is to be a modern, effective, working institution, can not forego the demands of its daily tenants for light, room, and air, and submit to the limitations set by calls for architectural effects, for imposing halls, charming vistas, and opportunities for decoration. The workshop, the factory, the office building, the modern business structure of almost any kind, these, rather than the palace, the temple, the cathedral, the memorial hall, or the mortuary pile, however grand, supply the examples in general accordance with which the modern book laboratory should be constructed (1897, 247).

As a piece of the adoption of a more businesslike architectural model, Dana recommends open shelving as a way to increase service efficiency – an argument that would undoubtedly have appealed to Carnegie and Bertram's corporate sensibilities.

Interestingly, within Bertram's "Notes," there is no explicit mention of either open shelving or free access to the books. However, this type of shelving is clearly assumed throughout, both in the provided plans and in the accompanying text. For example, Bertram describes the ideal small library as one "sub-divided as required by means of bookcases" and suggests that the reading-room windows should be higher than six feet off the ground "to give continuous wall space for shelving" (1916, 1-2). For the bookcases to be situated in these ways, the only reasonable access policy is an open one; it would be plainly impractical – and thus anathema to Bertram – to create artificial barriers between users and every inch of wall space just for the sake of limiting access to the books. Indeed, the lack of explicit

¹⁰⁴ Though Bertram does suggest that the selected site might be well-advised to contain space for a stack-room to be appended at a later time (1916, 2).

mention of closed vs. open stacks as an *issue* makes an interesting statement in itself: in essence, that the demands of “effectiv accommodation” had rendered the question moot; open shelving allowed for more efficient use of funds, and thus open shelving it would be.

Finally, it should also be noted that free access to library shelves would form an ideologically consistent extension of Carnegie’s views on both users and collections, as articulated above. Beyond the practical benefit of allowing for more idiosyncratic shelving arrangements before the standardization of classification systems, maintaining closed stacks in public libraries frequently came down to mistrust or even outright disdain for working-class library users. If users were allowed to roam freely through the stacks, it was thought, not only would the users be robbed of the librarian’s invaluable guidance, and thus select the “wrong” books, but the library’s collections would be thrown into disarray, and might even fall victim to theft or defacement (e.g., Ditzion 1947, 176). Contrast this with Carnegie’s views, described earlier, that the working man had all the potential in the world, and that given a range of books of widely varying quality and viewpoint, such a man would be able to discern between them, and use the “good” ones to improve himself. Indeed, Carnegie frequently exhorted the audiences of his library dedication speeches to read broadly and inclusively; to “Come with minds open, unprejudiced, hear all sides and hold fast to that only which is good” (Carnegie 1894). Such sentiments resonate strongly with the writings of later open-shelf advocates; for example, Dana suggests that the average library patron was coming to realize that “the only rational selection of reading is one made after the examination of many books; and [was] beginning to demand that he be permitted to come in immediate contact with the volumes he is invited to read” – and that any such system of immediate contact had to be based on faith and trust in the user:

The public library, whether it be a library which the people are taxed to maintain or a library which belongs to them by gift, must, so the demand goes, be managed with as much consideration for its patrons and with as much appearance of faith in their honesty as the ready-made-clothing house or the bookstore (Dana 1897, 245).

Although Carnegie never personally took any explicit position on the issue of open vs. closed stacks, the assumptions behind opening up shelf access seem well in line with Carnegie’s basically optimistic views regarding library patrons.

5. Conclusion

This chapter has provided an overview of the motivations behind and definitions of library use and collections within the Carnegie library program, as well as some of the relationships between those motivations and definitions and the policies and structures the program produced. Unlike in Boston, where the relationship between motivations, definitions, and implementation was essentially linear, the Carnegie program exhibits more multi-directional relationships between these elements. For example, Carnegie’s philosophy

of philanthropy and special love for libraries were infused to an ever-increasing degree within the program's implementation – its policies, conditions, and recommendations to communities. However, the same cannot be said for the definitions of users and collections present in the program; these, by contrast, exhibited very little sign of direct influence from the center of the program. Instead, those elements were largely decided within diffuse local communities – though the Carnegie program's stipulations helped to shape who was in a position to make those decisions and how.

In the name of efficiency, economy, and avoidance of philanthropic waste, intermingled with a single man's childhood love of reading, the Carnegie library program changed the shape of both information access and philanthropy itself across the United States and the world. This, to me, is an odd, if true, statement. After all, shouldn't information access programs be founded on high-minded ideals of democracy and education? Aren't public libraries supposed to be founded primarily to serve the public interest? While these elements certainly played into Carnegie's drive to spread libraries all over the English-speaking world, in this case they seem entirely subordinate to (a) Carnegie's desire to share his love of reading, and even moreso, (b) his desire to divest himself of a vast fortune in a relatively short period of time, without allowing his money to support indolence or laziness. Libraries met this criterion, while also supporting a pet cause – and \$56 million later, the Carnegie Library, in both mythos and physical form, had become an icon of public access to information.

Case 3: Google Books Library Project (2004-2012)

The Google Books Library Project (GBLP) is in many ways the keystone case of this dissertation; the project whose grand ambitions and public-private approach not only sparked my own interest in these issues, but also kindled the passions necessary to spur others into initiating several competing and complementary large-scale digitization projects that have followed in its wake – most notably, the Open Content Alliance, Europeana, HathiTrust, and the Digital Public Library of America. It is a case about which simultaneously a great deal and very little is publicly understood: a great deal about the legal issues and copyright lawsuit, but very little about the actual inner workings of the project: how it was conceived, by whom, and for what purpose(s) – and how those beginnings affected the shape it eventually took.

This chapter seeks to illuminate these still-dark corners of understanding, following the same outline established in the assessment of the two historical cases preceding it. As in those cases, this analysis will draw heavily on primary-source documentary evidence, including digitization contracts, press releases, terms of service statements, and FAQ pages. However, in contrast to the earlier cases, the early leaders of Google Books are still among the living. Thus, this chapter can – and will – draw substantially on first-person accounts from those individuals, gathered via semi-structured interviews conducted in person or over Skype. As noted in the earlier discussion of methodology, interviews were conducted with project leadership at each of the first six Google Books partner libraries – the University of Michigan (UM), Stanford, Harvard, Oxford, New York Public, and the University of California (UC) – as well as the Committee on Institutional Cooperation (CIC), which at the time of their agreement with Google was a consortium of 13 university libraries in the Midwestern United States.¹⁰⁵ Two interviews were also conducted with Google employees who had been involved in project leadership at high levels from the very early years. Some views from GBLP opponents interviewed for the next chapter, on the Open Content Alliance, will also be taken into account here.

It is worth noting at the outset that Google is a notoriously secretive organization, and that this secrecy had at least two points of impact on the research reported on here. First, very little additional documentary evidence was made available to me beyond what is already accessible to the public via the open web; virtually all such documents would most likely be covered under Google's proliferation of non-disclosure agreements (NDAs), as I was frequently told when I asked participants if they could share particular items that they had

¹⁰⁵ The CIC has since added an additional partner university – the University of Nebraska at Lincoln – but as they were not a member at the time of the contract (in 2007), I do not believe they are participating. Also, it is worth noting that the 13 libraries that were party to the agreement include two from the University of Illinois: one from its Urbana-Champaign campus and one from its Chicago campus (Committee on Institutional Cooperation and Google, section 1.3).

mentioned in their interviews.¹⁰⁶ As such, entire document genres that proved useful in the historical cases, such as correspondence and meeting minutes, were not available to draw upon here. Still, it should be noted that this is only partially due to secrecy: to a great extent, it also comes down to organizational procedures, and particularly non-documentation and deletion. For example, there is no indication that formal meeting minutes were ever taken in Google's meetings with its partners; indeed, it would seem out of keeping with the generally anti-bureaucratic Google ethos to have done so (see, e.g., Levy 2011). Most GBLP correspondence, in turn, took place via the inherently ephemeral medium of email, and much of it had undoubtedly been culled from the hard drives of both its sender and its recipient by the time data collection for this project began.

The second way in which Google's secrecy influenced data collection for this case – as well as the other interviews conducted for the dissertation – was in strengthening the case for participant confidentiality as a matter of project policy. Although I had no direct intention of digging into NDA-covered information in my interviews for this project, I also had no way of knowing what was covered and what was not (the parameters of the NDAs themselves not being public either). In case some such information leaked out into my dataset, confidentiality seemed a useful precautionary buffer against repercussions for informants.¹⁰⁷

Despite these limitations, however, the amount of data collected regarding this case was still immense – 12 interviews and 64 primary-source documents, triangulated against literally hundreds of secondary source accounts – and proved more than adequate to address the questions of motivation, definition, and implementation posed in this dissertation. As with the two previous cases, the focus here will be on the early years of the project: in this case, the period from about 2002-2007, when the definitions underlying the project were under most active negotiation, and its implementation in structures and processes remained in greater flux. Also as in the preceding chapters, a bit more historical background will be useful in order to frame the more substantive analysis that follows. Figure 35, below, provides a timeline of the project; a more complete version, linked to citations, has also been included as Appendix E. For clarity, different colors have been used to highlight three particular threads: additions of new library partners (blue), the copyright lawsuits (yellow), and growth in the number of volumes scanned (purple).

¹⁰⁶ At the very least, such documents ran the *risk* of being covered by NDA; another fact that became clear to me as the interviews proceeded was that the participants themselves were frequently quite unclear about what might be covered and what might not, and were understandably risk-averse about sharing given that uncertainty.

¹⁰⁷ This strategy for informant protection is further supported by the recommendations of several standard social research methodology texts (e.g., Krathwohl 1998, 215-16, Schutt 2009, 348-50).

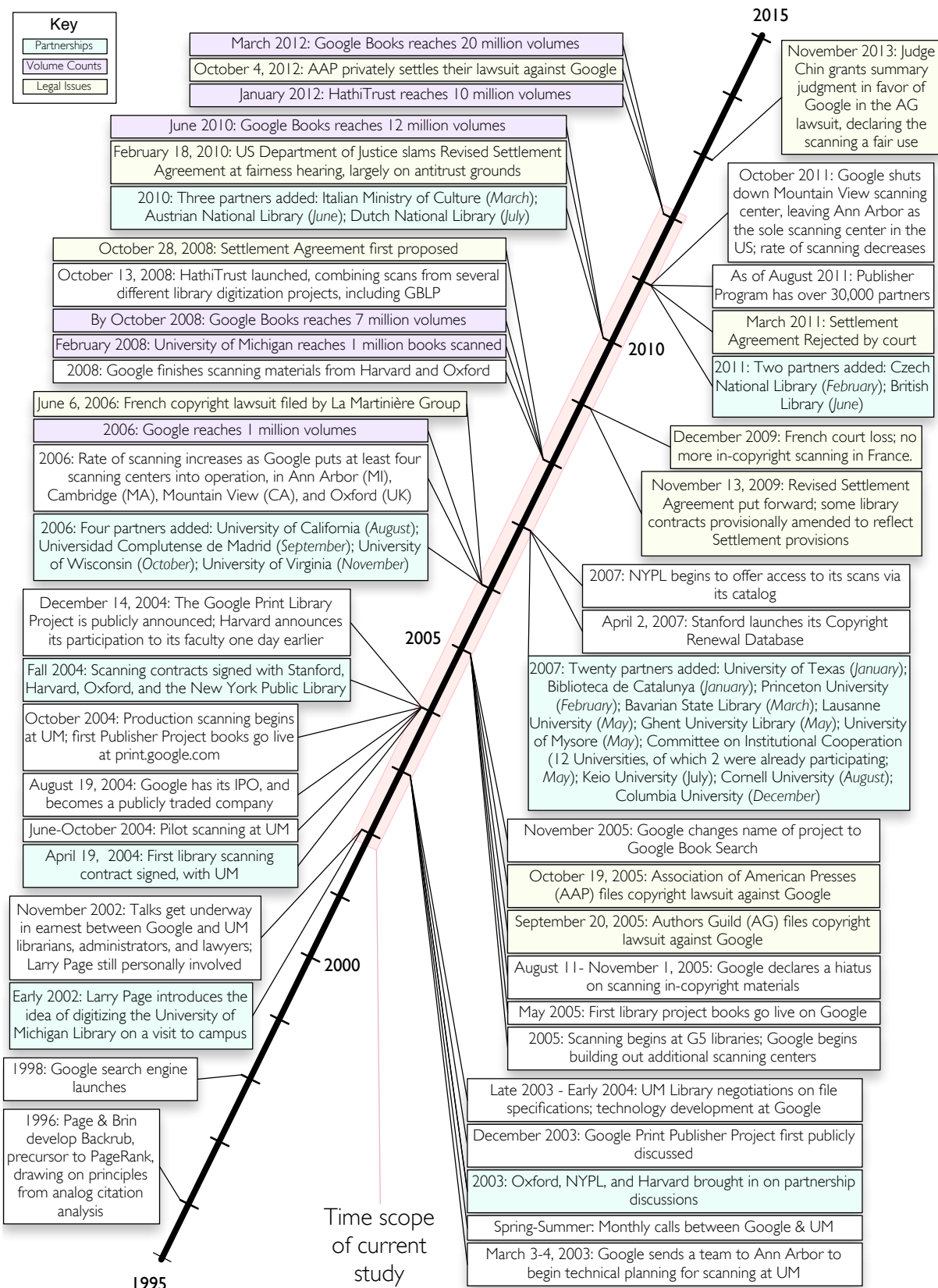


Figure 35: Timeline of Google Books Library Project (citations in Appendix E)

1. Historical Synopsis

The Google Books Library Project – or as it was then known, the Google Print Library Project – was announced to great fanfare and even greater controversy on December 14, 2004 (e.g., "Google Print Goes Live" 2004, Band 2004, Google 2004, Markoff and Wyatt 2004, Milliot and Albanese 2004, Quint 2004). However, the story actually begins several years earlier (a rough timeline appears in Figure 1, above). Indeed, many accounts date the genesis of the idea for Google Books to Larry Page's time at Stanford in the 1990s, before Google even existed. As the official Google history page for Google Books¹⁰⁸ reports, it was actually book citations that gave Page the idea for PageRank, the initial core of the Google search engine:

In the beginning, there was Google Books.

Well, not exactly. But one can certainly argue that the project is as old as Google itself. In 1996, Google co-founders Sergey Brin and Larry Page were graduate computer science students working on a research project supported by the Stanford Digital Library Technologies Project. Their goal was to make digital libraries work, and their big idea was as follows: in a future world in which vast collections of books are digitized, people would use a "web crawler" to index the books' content and analyze the connections between them, determining any given book's relevance and usefulness by tracking the number and quality of citations from other books.

The crawler they wound up building was called BackRub, and it was this modern twist on traditional citation analysis that inspired Google's PageRank algorithms – the core search technology that makes Google, well, Google (Google 2007).

In fact, Larry Page would later tell journalist Steven Levy that beyond being *inspired* by books, he and Sergey Brin had actually "tried to get book search going at Stanford." As Levy reports, however, the idea was not well received at that time:

"It would actually be really nice to be able to search all books," [Page would] say to his professors. "Why don't we do that?" It seemed obvious to him, but the professors deflected his suggestion. "They had other ideas about what that project [the Digital Library Initiative] was really about," he says. "If you asked anybody about it, they would immediately decide it was impossible" (2011, 347).

This tracing of the genesis of Google Books to before the dawn of Google itself is one recounted frequently, both in the media and in the interviews conducted for this study (P1, P4, P12, Markoff and Wyatt 2004, Quint 2004, Carr 2005b, Quint 2006a). And although one suspects it may be nearly as idealized as Andrew Carnegie's account of the formativeness of his childhood experiences with libraries, this genesis story, like that earlier one, has definite significance for the project's motivations; more on that in a bit.

¹⁰⁸ Oddly enough, this history of the project appears to have been abandoned in late 2007; it hasn't been updated since.

After Stanford, Larry Page's desire to scan and search the world's books next crops up in 2002. By that time, Google had gained a strong foothold in the search business, and was raking in enormous profits from its AdWords system (Levy 2011, 348). Technology had also progressed a great deal since 1996, making storage and transmission of the kind needed for such a project both cheaper and more powerful – and the trend was continuing in that direction. Page, by then co-founder and President of Products at Google, raised the idea of scanning all the world's books once again – and this time he had the money and the authority to make it happen himself. And in fact, it was Page, along with one of Google's early project managers, Marissa Mayer,¹⁰⁹ who actually scanned Google's first few books, using a metronome to keep time (Google 2007, Levy 2011, 349). These experiments in scanning, along with various cost scenario calculations, led Page to the conclusion that a project involving the wholesale non-destructive scanning of millions of books was, in fact, both technologically and financially feasible for his company. However, he still needed access to large quantities of books – and that was where the company's partnerships with libraries and publishers came in.¹¹⁰

The first library to be successfully approached was Page's undergraduate alma mater, the University of Michigan (UM).¹¹¹ On a visit to the UM College of Engineering in early 2002, Page casually mentioned the idea of scanning the entire UM library. Initially, those present weren't at all sure he was serious – after all, in 2002, scanning an entire 7 million volume library still seemed like an extravagant pipe dream. But he was in fact serious, and on November 7, 2002, University Librarian William Gosling briefed UM Provost Paul Courant on the idea; Courant was for it. A week later, on November 15, Gosling and Associate University Librarian John Wilkin met with Page to discuss the project in more detail. Shortly after, the University's legal department was looped in, and by March 2003, a team from Google was sent to Michigan to begin the technical planning for the project. Over the months that followed, Google and UM engaged in negotiations over file specifications, technology issues, and contract terms, culminating in the initial contract, agreeing to have Google scan the entirety of the UM Library's book collection, signed on April 19, 2004.¹¹² That contract in place, Google and UM worked together from June to October of 2004 to test

¹⁰⁹ Mayer went on to become President and CEO of Yahoo! in 2012.

¹¹⁰ The Google Books Publisher/Partner Program has its own complex history, which I will only briefly gloss here. For those wishing to know more about this aspect of the project, I would recommend the coverage in *Publisher's Weekly* (e.g., Milliot and Zeitchik 2004, Milliot 2007, Albanese 2010c, Albanese 2010d).

¹¹¹ As Levy notes, Brin, Page, and David Drummond (Google's General Counsel at the time) first approached the Library of Congress with the idea. However, although the Librarian of Congress, James Billington, was willing, the head of the U.S. Copyright Office (which is organizationally located within the Library of Congress), Marybeth Peters, was less enthusiastic, and put the brakes on the partnership (Levy 2011, 352); this opened the door for the group of five libraries that ultimately signed on to be a part of the project at its announcement in 2004.

¹¹² At this point the contract technically specified that this was a "pilot" agreement, from which either party could withdraw at any time without penalty.

the scanning technology and develop benchmarks for output quality. Production – that is, the actual contracted scanning of the books – then began in Ann Arbor in October of that year – still two months before the project was announced (P8, P10, P12, "MDP Timeline").

As all of this was happening at the University of Michigan, Google was also reaching out to other major libraries, as well as to publishers. The Google Print Partner Program – the early name for the publishers' side of Google Books – was first publicly discussed in December of 2003, and was officially launched October 5, 2004, with initial partners including Penguin, Hyperion, Scholastic, Houghton Mifflin, and several university presses ("Google Print" 2003, Albanese 2009b). Around this same time, Google began to engage in conversations with Stanford, Oxford, Harvard, and the New York Public Library (NYPL) about scanning books from their library collections (P13, P18, Carr 2005c). Still, none of the four would commit to signing a contract with the company until late 2004 (P8) – and this brings us back to the announcement of the project, with which this section began. That announcement, on December 14, 2004, was a highly orchestrated PR event, with all the "Google 5" ("G5") libraries agreeing to keep the project under wraps until that day. However, at the eleventh hour, Harvard University decided that its faculty ought to know first, and an announcement went out from their University Librarian the day before the official launch, on December 13 (Verba 2004). This occurred much to the consternation of the other participating institutions in the project, and foreshadowed the sort of behavior on Harvard's part that would leave several participating librarians with lingering ill feelings years after Harvard's participation in the project had ended.¹¹³

The power dynamics among Google and its initial five library partners were considerably less egalitarian and consistent than outward appearances made it seem at the time. By most interview accounts, the three most influential institutions were the University of Michigan, Stanford, and Harvard, in that order. NYPL and Oxford seem to have remained more at arm's length. During the period when the technical details were being worked out – how the scanning should happen, what the standards should be, and so forth – three individual librarians played especially influential roles: John Wilkin from the University of Michigan, Catherine Tierney from Stanford, and Dale Flecker from Harvard (P4, P8, P10, P12, P13, P17). These three collaborated a great deal with one another and with Google's library partnerships managers and engineers. Given the differences in approach among the three institutions, particularly in terms of conservatism vs. acceptance of risk, the triad was not

¹¹³ One librarian went so far as to declare that "Harvard of course [was] a source of great annoyance to almost all of us all the time. You know they're just so different, and you know they do things their own way, and they were very annoying as one of the five, one of the original Google five" (P9). This sentiment was echoed in several other interviews, at varying levels of tact and diplomacy. And lest it seem like this is just a matter of everybody backbiting everybody else on a complex collaborative project, I would note that almost every other mention of partnership dynamics within the GBLP was positive, and often extremely so; it was only Harvard that inspired this kind of bitterness.

always an easy one. As one librarian recalled,

kind of the word on the street... I don't know, maybe it was the word at Google, was that Michigan was kind of free and easy with what they let them do, and Harvard was proving to be very much a stick in the mud about what they would do and what they would send, and Stanford had this pleasant middle ground of being concerned, and careful, yet also kind of going with it (P17).

These differences in approach manifested themselves in the interpersonal negotiations among the librarians over the project's technical processes – and especially in debates between Flecker and Wilkin, who were seen as representing the two poles:

in the very early days, when there were just the five of us, and Dale Flecker was the Harvard lead for it, he had good technical knowledge so that he...and so he and John Wilkin could work against each – work off each other and come up with a good result. And that was valuable. So. In the very early days, Harvard did definitely have a role in some of these technical conversations. So again, sometimes they worked out conflict to the overall good, as I said you know Michigan is perceived as fast and loose, “yeah anything, we can do it”, and Harvard was not that way at all....So they had to, kind of... between the two of them, kind of come to a place that they were okay with. And they now and then agreed to disagree, which was fine, but they were both really good people in the middle of this (P17).

In interviews with librarians at these institutions conducted years later, those at Stanford and Michigan retained strongly positive feelings about the partnerships and their involvement in the project; Harvard less so. Indeed, under the leadership of Library Director Robert Darnton, Harvard has flung itself in the opposite direction from Google Books, becoming the initial institutional home of the Digital Public Library of America and standing firmly behind Darnton's series of essays opposing the Settlement Agreement.¹¹⁴ The individuals I spoke with from Oxford and NYPL were less vehement in their opinions about the project, but generally spoke about it in positive terms.

Contractually, each library's involvement in the project was different. The University of Michigan was the only G5 library to agree to comprehensive scanning of its collection, regardless of copyright status, to be executed purely in order of how convenient the books were to pull, on a shelf-by-shelf basis (P8, P10). Stanford also tentatively agreed to comprehensive scanning, and initially intended to pursue the same sort of “shelf-clearing” strategy as at Michigan, but ended up scanning more selectively, in order of legal risk (public domain first, then out of print) (P17, Markoff and Wyatt 2004). Harvard initially signed on to scan just 40,000 public domain books, but ended up scanning closer to 1

¹¹⁴ Hired in 2007, when Harvard was already almost finished scanning with Google, Darnton would likely never have signed on to the project in the first place; his opposition to the project has been vocal and frequent, especially with regard to the settlement. Darnton is also a leading figure in the Digital Public Library of America (2009, 2010a, 2010b, 2011a, 2011b, 2013).

million (P13, Markoff and Wyatt 2004). Oxford's Bodleian Library agreed to allow Google to scan everything in its collection published prior to 1900, but legal risk-aversion on Google's end eventually led them to shift that cutoff date back to the mid-1880s (P16, Young 2005a). And finally, the NYPL agreed to contribute between 10,000 and 100,000 public domain volumes, primarily of interest to scholars (Markoff and Wyatt 2004, Young 2005a).

As the project progressed, the contracts grew more standardized (e.g., University of Wisconsin and Google 2006, Committee on Institutional Cooperation and Google, University of Texas and Google 2007), even as the collections contributed grew more diverse: later agreements were structured with an eye toward filling in gaps in the Google corpus via the scanning of especially strong or distinguished collections at each institution, rather than scanning everything from one end of the library to the other. However, aside from the differences in the actual content to be scanned, later contracts would become virtually identical to one another (P12). Each library contract stipulated that the library would receive digital copies of all books digitized from its collection – although in some cases, libraries opted to allow Google to hold their digital copies of copyrighted works in escrow until the legal issues were more settled (P17, Committee on Institutional Cooperation and Google). All contracts also contained limitations on what the libraries could do with their digital copies, although none of the Google partner librarians I spoke with seemed to find these restrictions especially onerous (though some did wish there were fewer of them). Among the public contracts, the restrictions seem to be geared mainly toward preventing competition with services Google is offering (or facilitation of such competition by others), largely by limiting automated access to the files – that is, restricting access by search bots and preventing bulk downloads (University of Michigan and Google 2004, University of California and Google 2006, University of Wisconsin and Google 2006, Committee on Institutional Cooperation and Google, University of Texas and Google 2007).

In early 2005, Google set about building scanning centers near the G5 libraries, and began to scan (P4, P16); the first library project books would go live on Google that May, joining Partner Program books already online (Last 2007). Alongside this progress, a steady drumbeat of opposition to the project was growing among publishers and authors, many of whom viewed the scanning and indexing of in-copyright works as a violation of their intellectual property rights, and often declared as much using wildly hyperbolic rhetoric – averring, for example, that Google was “Not only...trying to rewrite copyright law, it is also crushing creativity” (Schroeder and Barr 2005), or, more bluntly, that “Google [was] taking our property and not paying for it. It's burglary” (Graham 2005, quoting Nick Taylor of the Author's Guild). In response to these critiques, Google simultaneously altered the project's opt-out policy to allow copyright holders to opt out of having their work scanned at all (where before it had only allowed them to opt out of short snippets of the work being made available on the Google website), and placed a three-month hiatus on the scanning of in-

copyright works in general, from August to November of 2005 (Albanese 2005b). However, these concessions were insufficient to placate irate copyright holders, and within that three-month window, two major copyright lawsuits were filed against Google over its Books project: the first by the Authors Guild (AG), on September 20, and the second by the Association of American Publishers (AAP), on October 19 (Helm 2005a). These lawsuits were soon joined into a single case, and remained so for several years.

The lawsuits notwithstanding, Google and the libraries resumed scanning of in-copyright works in November of 2005, and shortly afterward the name of the project changed from “Google Print” to “Google Book Search,” more accurately reflecting the initiative’s aims – to index and search books, and not to sell or distribute anything physical (Gohring 2005, Helm 2005a). The next two years saw the program’s expansion: 24 more library partners were added,¹¹⁵ the rate of scanning increased, and the project hit its first major scanning milestone, of 1 million volumes (P4, P12). Additionally, a major policy change went into effect, greatly broadening effective access to Google’s public domain scans: that is, in September 2006, the company began to allow free PDF downloads of all public domain works scanned from libraries, and also began to place “Find in a Library” links alongside purchasing links in each book’s left navigation (Albanese 2006a, Quint 2006b).

During this time, there were at least three scanning centers in operation in the United States – in Ann Arbor, Michigan; Mountain View, California; and Cambridge, Massachusetts – as well as one in the UK, at Oxford (P8, P12, P16, P17).¹¹⁶ As these operations got going, the project leadership at Google quickly realized that a huge part of the challenge involved in scanning whole libraries full of books derived from the pure logistics of the task: how would the books get from the shelves to the loading docks to the trucks to the scanning centers and back? And how could one make that happen as quickly as possible? To help address these issues in a systematic way, the company hired a logistician, Doug Kuch, in 2006 – by all accounts, an extremely important decision, as Kuch’s involvement greatly improved the project’s efficiency and flow (P4, P8, P12).

Google’s scanning partnerships with at least two of its five original library partners – Harvard and Oxford – concluded in 2008 (P13, P16).¹¹⁷ As of 2013, the University of

¹¹⁵ The partners added in 2006-2007, in chronological order, were the University of California, Universidad Complutense de Madrid, the University of Wisconsin, the University of Virginia, the University of Texas, Biblioteca de Catalunya, Princeton University, the Bavarian State Library, Lausanne, Ghent, Mysore, the Committee on Institutional Cooperation (12 university libraries, of which two were already participating), Keio University, Cornell University, and Columbia University (Colvin 2006, University of Virginia Library 2006, University of Wisconsin 2006, Crawford 2007, Grogg and Ashmore 2007, Biblioteca Complutense).

¹¹⁶ There may well have been more; however, those are the four for which I have positive confirmation.

¹¹⁷ It seems likely that the New York Public Library also finished scanning with Google around this time, although I have not found direct substantiation for this assertion (in any case, NYPL had finished well before 2012, when I conducted my interviews).

Michigan, Stanford University, and several others continued to scan with Google, though at a much slower rate since the settlement agreement was rejected in 2011 (Howard 2012). Among my interviewees at partner libraries, the failure of the settlement was broadly perceived as a major turning point in the project: the point at which the wind went out of Google's proverbial sails, and the company's enthusiasm for the project – already having fallen victim to a certain level of bureaucratization – truly began to spiral downward (P3, P8, P9, P17).¹¹⁸ In those interviews, the settlement tended to be discussed with a sort of wistfulness for what might have been, and specifically for what it might have allowed libraries to do: in particular, greater freedom for the partner libraries to make alternate uses of their public domain materials, and greater access for everyone to the in-copyright materials scanned by Google. As Andrew Albanese reported in 2008, the settlement would have provided several benefits for educational users, including:

- Free preview and ability to either find the book at a local library or through a consumer purchase
- Creation of a database of both in-copyright and out-of-copyright (public domain) works on which scholars can conduct research
- Institutional subscriptions providing to in-copyright, out-of-print books
- Accommodated services for persons with print disabilities
- Digital copies of works digitized by Google provided to the partner libraries for “long-term preservation purposes”
- Free full text access at public libraries around the country (at a single designated terminal) (Albanese 2008c)

Beyond this, of course, the settlement also would have had major benefits for Google and the publisher and author groups involved in the case, by facilitating a new market for out-of-print, in-copyright books, funneled exclusively through Google, with the publishers and authors also receiving a one-time \$125 million payout from the company to settle the case (Albanese 2009b).

However, the exclusivity of this agreement, combined with its truly massive implications for the future of access to digital books, ended up making the settlement unpalatable to many stakeholders, including some who had been strong proponents of the underlying

¹¹⁸ The first version of the Google Books Settlement was proposed in October 2008, with a second, revised version submitted in November 2009. Ultimately, however, the revised settlement was rejected by the court on copyright and antitrust grounds in March 2011 (*Rejected Settlement Opinion* 2011, Albanese 2011b). Both versions of the Settlement Agreement received a huge amount of critique as well as a fair bit of support; *The Public Index* has assembled a relatively comprehensive collection of this documentation, including legal filings, news coverage, and analysis related to the lawsuits and settlement agreements (<http://thepublicindex.org/>). Ultimately, the AAP reached a private settlement with Google in October 2012 (Sporkin 2012a), and the AG's lawsuit was decided in Google's favor on summary judgment in November 2013 (Chin 2013); the AG's appeal in that case is pending as of the time of this writing.

scanning (e.g., Samuelson 2009, Lessig 2010).¹¹⁹ The exclusivity issue is well-summarized by legal scholar James Grimmelmann:

The settlement would let Google sell out-of-print books unless the copyright owners objected. A competitor, however, would need to get individual permission first or be sued into oblivion. That's hard enough in general, and for orphan books it's impossible. There's no one to ask. The class action opens a door for Google, but leaves it closed for everyone else. That fact has always been the absolutely critical feature of the settlement. It makes the settlement exciting, because it means that millions of out-of-print and orphan books would become much more widely available. It also makes the settlement dangerous, because millions of copyrights would be collectively press-ganged into Google's service (2009b).

Ultimately, of course, the settlement *was* rejected (although not until 2011, following several long delays), and since then the future of the project has become much less certain. Although the AAP would eventually reach a private agreement with Google, separate from the AG, its terms remain confidential (Sporkin 2012a). Both the AG lawsuit against Google and the organization's parallel lawsuit against HathiTrust were decided on summary judgment in 2013 and 2012, respectively, with both judges finding that the scanning and uses being made of the products of the scanning constituted fair use under U.S. copyright law; both of these lawsuits remain under appeal as of mid-2014 (Albanese 2012, Mullin 2013, Smith 2013). No additional library partners have been added since 2011, and by October of that year, Google had shut down all but one of its U.S. scanning centers – its first one, in Ann Arbor, Michigan – which means that now all books from even far-flung U.S. institutions like Stanford must be shipped cross-country for scanning (P17). As one librarian put it, prior to 2011, the libraries and Google “*were sort of on a train trying to get to where they’re going as fast as they could. And, you know, now it’s like, ‘oh! We’re stopping at every station!’*” (P9). Still, the train continues to move, if much more slowly. At last count, the project has led to the digitization of more than 20 million books, the vast majority of which have been scanned from library collections,¹²⁰ and the number grows larger every day (P12, Howard 2012).

¹¹⁹ This category also includes me: I continue to believe that large-scale digitization of the sort pursued by Google represents a fairly massive net benefit to the cause of access to information in the world – and that the underlying scanning should represent a fair use under copyright law – but I also believe the settlement was a massive overreach. My armchair guess at what might have led to that overreach? The parties got behind closed doors, and started to dream big. Too big. They saw all the good they could do, and forgot to consider the fairly glaring oversights they were making with regard to things like revenue distribution and industry consolidation. Perhaps this is an over-generous assessment – perhaps it was a more cynical grab for money and control on some or all sides – but given my interactions with several of those involved, and their enormous enthusiasm for the cause of opening access to books, I do not think it radically so.

¹²⁰ It is also worth noting, however, that at this point Google has signed up virtually every publisher of any significance for its never-controversial Partner Program – at last count, more than 30,000 of them (Albanese 2011a).

2. Motivations

The creation of the Google Books Library Project reflected a complex web of motivations which differed significantly between Google itself on the one hand, and the company's early library partners on the other. This section will discuss first what is ascertainable about Google's motivations,¹²¹ and then move on to the libraries' aims. In particular, although both Google and the libraries link the book scanning project to their respective institutional missions, and evidence also points to utilitarian interests on both sides, the specific shape and manifestation of both of these types of aims differ greatly between Google and the libraries. On the one hand, available evidence suggests that Google at least initially saw book-scanning as a labor of love, an interesting technical challenge to be conquered, and a way of proving that such things were possible. On the other, the libraries took a broader view, expressing hopes that the project would change the world, and in particular, that it would usher in a more modern – or even futuristic – form of library, that it would make possible previously unimaginable uses of information, and, ideally, that it would act as a catalyst for positive change in the legal regime surrounding books and reading. This section will discuss each of these groups of motivations in turn: first Google's, then the libraries'.

2.1 Google's Motivations

The nature of Google's interest in the project has changed a great deal over time. At the outset, the project was driven by a heady combination of personal interest and company mission – a passion project of one of Google's founders, perceived as falling closely in line with Google's official mission to “Organize the world's information and make it universally accessible and useful” (e.g., Langley and Bloomberg 2006), initially overseen by an enthusiastic team of Googlers¹²² who believed strongly in their project's potential for social benefit. Later on, as this early staff moved on to other divisions or other companies, the project became more corporatized and impersonal, and seems to have relied more heavily on pragmatic, utilitarian motivations to propel it forward. As one partner librarian suggested, by 2011, it had “*become more of a product, less of a...mission*” (P8). The early mission-driven aims are well reflected in both the interview and documentary data; the later utilitarian aims less so – in part because any discussion of the pragmatic goals of a corporate project quickly begins to bump up against strategic (and thus secret)

¹²¹ As noted, Google is a very secretive company. Although the people I spoke with there were extremely helpful and forthcoming in their interviews, their responses mainly spoke to their own experiences and understandings; corporate strategy was decidedly not on the table as a topic of discussion. Indeed, Google was never even especially forthcoming with its library partners about its motives. As one librarian mused, “*There's some pretty clever people on top of that company. They were never very clear. And they would be very honest with us and say 'look, we're just not going to tell you what our business plan is. We don't have to. Because we're Google, and you're not.' That's OK. [laughs]*” (P3).

¹²² This is how Google staff refer to themselves; I've adopted the term here for convenience (e.g., Levy 2011, Google n.d.-b)

information.¹²³ In this subsection, I will discuss (1) the motivating role of mission and mythos, (2) the question of profit motive, and (3) the extent to which the project may have been motivated by a desire for ever-greater quantities of data to crawl.

2.1.1 Mythos and Mission

As noted in the historical synopsis above, when Google Books genesis stories are told, they almost always trace the project's origins back to a pre-Google Larry Page pondering book citations at Stanford, and build upon that historical anecdote to argue that Google Books is (a) as natural and rational an activity for Google as any could be, and (b) a necessary piece of Google's mission to organize the world's information. Indeed, when asked directly about the motivations behind the project in an interview, one Googler immediately responded by recounting exactly this origin story, weaving it seamlessly into the company mission at the end:

Well, you know, um... I don't know if you know the history of how Google started with Larry and Sergey at Stanford... but you know, the original Google project was, um... was I think was called the Digital Libraries project. That was the, um... the original project that Larry and Sergey did where they...which resulted in the PageRank algorithm for...Google. The...algorithm that made Google such a successful search engine. Sort of started as a thought exercise in the library. They had this idea that well if you could take all the books in the library, and you could organize them, not just by subject and so on but if you could link them together by citation, you know, like this volume refers to this volume, and this volume refers to this volume, and then you did sort of a hierarchical ranking based on the influence of one idea and how it sort of percolated through, through other volumes...what kind of ranking would that be? You know what would that look like? But to do something like that you'd have to digitize all the books, and they hadn't digitized all the books, so... they took that idea and they applied it to the web. Thinking of hyperlinks as being sort of like citations. And doing sort of this complicated citation analysis of the web which turned into PageRank. So the founders were thinking about this sort of thing for a long time. I mean, I remember anecdotally, um...back in 2002 or 2003 hearing some meeting, or some...I think it was a just a general public conversation where Larry and Sergey were talking about scanning books. So I know they'd been thinking about it for a long time before...before we actually did any of it. And I guess...I guess that once we got pretty good with, um... with the web, and had done a decent job organizing it with the Google Search engine they turned their sights to information outside the digital world. Or information that wasn't born digital. I mean if you think about it, the web is great, but everything on the web is sort of by definition 30 years old, pretty much? Unless it's been digitized. Or copied from older material. And most of human knowledge is more than 30 years old. And most of it is not recorded on the web. So... it was sort of a natural extension of the...of the "organize the world's information and make it universally accessible and useful" mission (P12).

¹²³ It is worth noting that there is, in fact, significantly less data about Google's motivations in my dataset than there is about the libraries'. For scale, among 208 primary source excerpts for this case coded as dealing with motivations in some way, 166 discussed libraries' motivations, vs. 48 for Google's (with some overlap between them).

This anecdote serves to naturalize and personalize the project, and also to provide a logical, non-cynical (i.e. non-profit-related) way of understanding what Google is doing in scanning books. In this, and in its content, it has two key similarities to the parallel mythos attached to the Carnegie library project. First, it makes the project personal rather than business-related, which makes it feel less threatening. Just as many harbored fear and anger toward Carnegie Steel in the late nineteenth century, many harbor fear and anger toward Google in the early 21st. Easily thousands of articles and web postings – even a few books – have been devoted to explicating the threats Google and its dominance pose in myriad areas, including (but hardly limited to) copyright, privacy, and freedom of expression (Cleland and Brodsky 2011, Vaidhyanathan 2011; also several now-defunct or offline sites like Daniel Brandt's Google-Watch.org and Phillip Lenssen's Google Blogoscoped; many additional sources provided on the Wikipedia page for "Criticism of Google"). And as with Carnegie a century earlier, many of these critiques (though decidedly not all) are both valid and well worth raising. In terms of rhetoric, then, distancing the Books project from the evil-corporate-specter end of Google, as a personal dream of one of the founders, helps to make the initiative seem more sympathetic than it might in a less personalized form.

The second key similarity between these two origin stories lies in their mythologization through repetition. Like the story about Carnegie being profoundly influenced by his time in Colonel Anderson's library, the story about Larry Page being inspired by citation-linking in books, and wanting to scan them all to make that happen, seems plausible enough. However, also like Carnegie's, Page's story has been told and retold so many times that it is likely that all the rough edges of reality have been smoothed out; all the other factors that might have played in, all the complications that fell into the path – they have all been progressively brushed aside over a long succession of retellings, leaving behind a more smoothly flowing narrative and rhetorical arc. So has Larry Page been interested in scanning books since before Google existed? I see no reason to doubt this, just as I see no reason to doubt the basic outlines of Carnegie's interest in book provision. Still, to imagine that the smooth-lined canonical version expresses the full scope of either man's motivations would be naïve in either case.

2.1.1.1 Google's Mission

Whatever the veracity of the Larry Page mythos, however, there is little doubt that a sense of mission – and particularly Google's corporate mission, to "Organize the world's information and make it universally accessible and useful" – played a strong role in propelling the project forward in its early years. There is a significant extent to which, simply on its face, scanning books – and even scanning *all* the books – really does simply seem like an extension of Google's mission: after all, books are part of "the world's information." By scanning books, Google is making the information within them more "accessible and useful," to a vastly greater proportion of the world's people (if not fully

“universally”). As Google General Counsel and Vice President¹²⁴ David Drummond suggested just after the filing of the copyright lawsuits,

this program, which will make millions of books easier for everyone in the world to find, is crucial to our company's mission. We're dedicated to helping the world find information, and there's too much information in books that cannot yet be found online. We think you should be able to search through every word of every book ever written, and come away with a list of relevant books to buy or find at your local library. We aim to make that happen, but to do so we'll need to build and maintain an index containing all this information (2005).

And further, for the staff working on Google Books in the early years, I would suggest that the term “mission” carried more than the usual corporate connotations: when interviewees (both from libraries and from Google) talked with me about mission as a motivation at Google at that time, there was an element of *calling* about it; an almost missionary outlook. There was a sense, in short, that the early Google Books team genuinely saw themselves as providing a service of benefit to the world, and were personally invested in the execution of that service, in a way that went well beyond serving the needs of their company or its business priorities.

Several interviewees, both at libraries and at Google, expressed the view that mission was *the* primary motivator on the Google side, at least in the years before the settlement. Certainly mission appeared to be a much greater motivating factor than a desire for profits. As one library partner pointed out, “*if you’re already the A-number-one-primo biggest kid on the block, do you really need to spend several hundred million dollars to get another clump of relatively esoteric information? Sure, you can build n-grams, but is there really any money in building n-grams?...At the time that it started it felt like mission*” (P10). And interviews at Google echoed this assessment. One Googler I interviewed explained at length how leaving books *out* of Google search would feel, to him, like leaving them obscure to the world – the antithesis of Google’s stated mission, and a failing about which he would feel personal regret:

The vision was that there was a vast amount of information, and that our...mission was to...help organize it and...make discoverable information and not just webpages, and that all this information, which was the way that, um, society had been recording its insights, its, its poetry, its creativity, its, everything that is- was through these artifacts, and in the digital age, these artifacts were um, uh, at some level being left behind. And, you could search the World Wide Web, but and, and that, and that full text search...had radically shifted our ability to discover...information. And the way it used to work in a library, would be, you’d go at...the topic in a card catalogue, and...you look at the title, and you really have no idea if what you’re looking for is in there. And full text search on the Web radically shifted...people’s ability to find information and changed the way we, um...our efficiency in finding stuff and consuming stuff and finding the knowledge we wanted.

¹²⁴ That is, at the time. Drummond is now Google’s Senior Vice President for Corporate Development and Chief Legal Officer (<http://www.google.com/about/company/facts/management/>).

And the idea that this – this content would be left behind...in the digital age, um, I think is a, um, y'know, sad concept (P4).

From there, this participant went on to imagine how differently his son might use books once he entered college in 2025, and to muse upon the immensely expanded set of digital resources that would be available to him once that time came because of projects like Google Books. Ultimately, he went so far as to declare that his work on Google Books was the proudest achievement of his career; the thing that he hoped “*forty or fifty years from now my kids or my grandkids...will y'know know that daddy did*” (P4). The other Googler I spoke with shared this enthusiasm, though from a different perspective, explaining that, “*the reason I went to library school was to help people find the information they were looking for. And I feel like this project... I'm doing this maybe one step removed from the actual patron, but, um... still have a lot of impact on people finding what they need to find out*” (P12). Certainly, there is an element of in-group conformity that comes through in these individuals' expressions of personal passion for the project, but the underlying sentiments nonetheless seemed quite genuine.

2.1.2 Profit motive (or not)

As noted above, a desire for profits does not appear to have been a strong initial motivation for Google Books – at least not for the Library Project, and not in the early years. For one thing, scanning books was and is still relatively expensive (at least, when pursued by the millions),¹²⁵ and recouping that expense no small matter – much less going beyond cost recovery to actually profit from the endeavor. Certainly, book sales would not soon accomplish this aim – especially given the proportion of orphan works in the corpus, for which no permission to sell could be sought or given (e.g., Wilkin 2011), the general obscurity of the contents of most academic library collections,¹²⁶ and the overall downward trajectory of the publishing industry over the past few decades. Of course, book sales are not the only feasible source of income from this project; the data it adds is of far greater value, and a much more likely candidate for propelling Google's profit strategy for its Books project (on that, more in the next section). But still, as one Googler notes, “*it's not like [book digitization] is the area where there's big business,*” although in many years, it might

¹²⁵ Brewster Kahle estimated the cost of Google's scanning at \$5-10 per book (2009c), which would place the cost of scanning Google's corpus of 20 million books somewhere between \$100 million and \$200 million. For scale, Microsoft, by far the largest single financial contributor to the Open Content Alliance, spent just \$10 million before concluding the project was not strategically advantageous – less than 10% of Google's likely *minimum* expenditure to date, for the scanning alone (Guess 2008). Indeed, it should be noted that even these enormous figures most likely do not account for many elements of the actual costs of scanning, such as the cost of building the technology in the first place, or other infrastructural or logistical costs as time went on – and they certainly do not account for the legal fees, which must be immense.

¹²⁶ As Michigan's University Librarian, Paul Courant, is fond of saying, “most of the works in question have been out of print for a reason” (Courant 2008, comment on November 9, 2008).

“actually one day really pan out, to where it is economically feasible” (P4). In pure ROI terms, investing in the scanning of millions upon millions of obscure academic texts seems a fairly perverse way to go about maximizing profits for a company like Google. (Indeed, this is essentially the conclusion that Microsoft reached from its short-lived foray into this area with the Open Content Alliance; the announcement of the project’s end notes the company’s decision to focus more “on verticals with high commercial intent,” such as travel (Nadella 2008)).

In my interviews, profits were never once cited as a motivation for Google Books by Googlers. They were mentioned as such by some librarians – but not necessarily in the affirmative. One later-entering partner librarian did express a strong belief that there must be a central profit motive behind the project (P3). However, the G5 librarians I spoke with universally dismissed the idea that money was a major consideration in the earliest days. As one explained:

for Google, I take them at their word. I don’t think they had a business plan, I don’t think they had any notion about selling books or getting on or whatever. I think they just thought “a lot of words. A lot of search terms. A lot of credible content” to go along with this... what back then would have been much more fledgling, dubious web content. So, you know, I think it was just a way to improve, improve search. And get context to it. So, uh... You know I just don’t think they had a clue. I don’t think they had a clue as to what it would cost, or what they’d be getting themselves in for, and I don’t think they had a clue as to what they were going to do with it all. But, you know, probably figured they would somehow... you know there would be value there. So... that’s probably naïve, but that’s what I think about Google (P9).

Another librarian echoes this assessment, reflecting: *“I’ve never thought... So I think Google was delighted to learn that under the settlement agreement they might have made a lot of money. You know, if you’re in business making money is okay. But this was never really about the money”* (P10). Although the profit motive increased in importance over time, as the project made its aforementioned transition from passion project to standard Google business product (P8), literally nobody I spoke with who was involved in the very earliest years of the project was willing to ascribe Google’s initial interest in books to a desire for profit in any truly significant way.¹²⁷

¹²⁷ Furthermore, this assessment meshes well with other accounts of Google’s company-wide subordination of profit considerations to their interest in making useful things. One of Levy’s anecdotes in the preface to *In the Plex*, from a trip he took with Marissa Mayer and several Google Assistant Project Managers to Bangalore, provides a near-ideal illustration of this product- and user-oriented culture: “An interesting moment,” he writes, “occurred in Bangalore when Mayer was taking questions from local engineers after presenting an overview of upcoming products. One of them asked, ‘We’ve heard the road map for products, what’s the road map for revenues?’ She almost bit his head off. ‘That’s *not* the way to think,’ she said. ‘We are focused on our *users*. If we make them happy, we will have revenues’” (Levy 2011, 5).

In fact, the most frequent and insistent attributions of a profit motive to Google in this case came not from participants in the Google Books project, but from participants interviewed regarding their work on the Open Content Alliance – many of whom have been outspoken critics of Google and/or Google Books from the very beginning. For example, one opined:

You know Google is a commercial project that is...that was created truly to drive that little search box. You know? It was content that could be used to drive the search engine....So...You know, it's definitely nice to be able to use the Google Books content, I use it myself. It's nice to be able to be up at 11 o'clock searching through, y'know, 10 million books, but its spirit was totally different. It's meant as a commercial...product (P11).

Another OCA interviewee, though not opposed to the Google project, nonetheless drew a profit-based distinction between that initiative and the OCA, explaining that:

Google doesn't... simply doesn't have the same, uh, DNA, as a memory institution does. You know, it doesn't have the same requirements or expectations, or... um... ah... MISSION. You know? It's there to make money. And it's... basically it's an advertising company that just has a shitload of data (P7).

The fact that these sorts of assertions came exclusively from non-participants in the project, who had little or no firsthand interaction with the Google team involved, while those who did have such interactions declare exactly the opposite, seems to me quite telling. In the absence of actual information, human beings are excellent at ascribing rationales that fit with their own existing worldviews and agendas; I would suggest that there is some such rationalization going on here on the part of those not directly involved.

On the other hand, to say that profits were not a *primary, early* motivation for the project is not to say that they were not a motivation at all. Indeed, if the company was not seeking to monetize this initiative in some way, it would be defaulting on its responsibility to its shareholders as a public company. And beyond this, Google's secrecy about the project – and particularly about the scanning technologies it developed for it – as well as the limitations it has placed on what libraries and users can do with the scans, are indicative of a desire to maintain some level of competitive advantage with regard to book scanning – if there is money to be found in large-scale digitization, they would like the opportunity to find it first (indeed, this connection is made explicitly by a few interviewees, especially P4, P10 – although the latter tends to dismiss the idea that any of Google's competitors would have had any serious interest in pursuing scanning at this scale themselves). But even so, as a few other interviewees (both participating in the project and not) suggested, it seems likely that Google's profit strategy here had much more to do with hoovering up data than with selling books (especially P2, but also P3, P1, P7, and P11). After all, Google's massive stores of data are what run their ad business, which is what produces the vast majority of their revenue. More content data (like that in books) brings in more users, who produce

more user interaction data (like choosing, buying, reading, and annotating books), which increases advertising revenue both by broadening the audience for Google products and by providing the raw materials for improvements to search and ad-targeting algorithms.¹²⁸ As Andy Rubin, then Google's Director of Mobile Platforms, once bluntly declared, "We don't monetize the thing we create....We monetize the people that use it. The more people that use our products, the more opportunity we have to advertise to them" (Levy 2011, 229). By offering the opportunity to search the full text of books for free (though, importantly, not necessarily to actually *view the text* of those books), Google can gather useful data for improving its algorithms, targeting its advertising, and bringing more eyeballs to its family of websites – thus increasing profits, but in a less direct (and likely more lucrative) way than selling books directly to customers.¹²⁹

2.1.3 Get More Data

As noted above, Google's drive to collect ever-larger heaps of data, including those in and surrounding books, is partially about making more money. Such data, after all, form the core of Google's business. At the same time, however, the data-collection element of Google's book-scanning motivations also partially reflects what I'll call the "this is cool" factor: the desire to get more data just to see what sorts of things clever engineers might be able to do with it. As one library partner mused when asked about Google's motivations,

Engineering. Just engineering problems are so interesting. With...these guys...you know, for them. Mass digitization: there's some really interesting engineering search problems that you can, you know, if your – one of your primary things is search, imagine how much you can improve on search if you have this enormous corpus of material in multiple languages. From published works. So, you know, it's a research corpus, I'm sure (P3).

Another partner librarian concurs, linking Google's desire for data in this case to both the usefulness of digitized book contents for improving artificial intelligence and the advertising revenue to be realized both from users of the digital books and from more precise algorithms and ad targeting throughout the Google product sphere:

I deeply believe from Google's perspective that their primary benefit, um, in digitizing this material, is to gain a huge amount of data. Um... via the digitization of these resources... and that data can be broken down in a lot of different ways, um, I mean there's some examples that are pretty obvious, um, there's a lot of latent semantic analysis that can be done with these texts, which would obviously optimize search

¹²⁸ Indeed, according to Levy, this is precisely the argument that won Google CEO Eric Schmidt's support for the project. As he reports: "'Eric wasn't skeptical but listening, trying to make sense,' says Megan Smith, the biz-dev person who became involved in the project. 'If something passed his directional sniff test, if there was a business reason behind an idea, he was open to things.' In this case, Schmidt became convinced that capturing books in Google's search index would allow Google to deliver important information that was currently lacking – and that eventually the investment would be recovered by increased traffic and more clicking on ads" (Levy 2011, 350).

¹²⁹ Though they would ultimately begin to provide that service as well, as of December 2010 (Albanese 2010d).

queries... ah, there's a tremendous amount of knowledge, um, locked within, um, nonfiction works, that could be extracted... A lot of entity relationships that could be determined... um... the fact that works are available often in multiple languages means that translation capabilities can be, um, enhanced with far more sophistication than, uh, other computing-intensive, ah, analytical approaches would be able to garner on their own.... Um, and so forth. Um, and so I think in terms of the potential advertising revenue that Google would be able to garner through digitization, uh, this kind of project's very attractive (P2).

Overall, the desire to extract all this data from books – and just to have and use that data – seems likely to have been an extremely compelling motivation behind the project. However, I have no confirmation of this from the Googlers I spoke with – likely because this is one of those areas (actual business strategy) where Google's NDAs would come into play. The librarians, however, were free to speculate, and given their long relationships with the company, I give their assessments a fair bit of credence.

Moreover, the “this is cool” element of getting more data meshes quite well with the Larry Page mythos described at the beginning of this section: Page simply seems to have been extremely personally enthusiastic about this endeavor. As Levy recounts:

Page was rhapsodic when explaining the [initial deal with the G5 libraries]. At Stanford, he said, he had heard there were 132 miles of books in the libraries, but you couldn't find what was in them. Google's project might drive people to go to libraries more often, because now they would know what was in there. 'That's the really big deal,' he said. 'A lot of people thought that this was impossible.' (Levy 2011, 357).

Certainly, there is still much that is obscure about Google's motivations for scanning the millions of books now in its databases. However, between one founder's enthusiasm, the passionate pursuit of company's mission among its staff, and the technological and monetary benefits to be reaped from the massive piles of data contained in books (and tracked around the Web by their readers), one can construct a reasonably adequate picture of the elements that spurred the GBLP into existence – and so I have attempted to do here.

2.2 Libraries' Motivations

Each of the libraries that partnered with Google on its book scanning project did so for their own particular set of reasons, in service of the unique set of ideological and pragmatic commitments held by their institution. However, among this diverse set of expressed motivations, there exist strong continuities, which can be grouped into two rough categories not unlike those used for Google: mission and pragmatism. For libraries, the mission-oriented rationales for participating in Google Books tend to fall into two basic parts – access and preservation – though the precise scope of these terms varies across institutions. The pragmatic motivations for participation are more diverse, but can be grouped roughly into two groups as well – a desire to improve and expand existing library

roles (digitization-related and otherwise) and a sense that the project would both enhance the reputation of the individual participating institutions and help to ensure their collective ongoing relevance within a rapidly changing information ecosystem.

2.2.1 Mission

In public statements announcing and explaining their digitization partnerships with Google, staff at GBLP partner institutions have made frequent and direct reference to institutional missions, both of the libraries themselves and of the universities that house them.¹³⁰ These statements especially focus on libraries' twin missions of access to and preservation of information resources, which are generally mentioned in the same breath (e.g., Coleman 2006, Colvin 2006, Stanford University Libraries 2006, Barnett and Willmann 2007, Bibliothèque Cantonale et Universitaire – Lausanne, University of Michigan). A passage from a Columbia University Library press release is typical of such statements:

As a research library, our mission is to collect material of value to teaching and scholarship, to connect people to the information in those resources, and to ensure access and preservation of our collections into the future. We believe that our partnership with Google will significantly advance that mission (2007b).

The word “mission,” couched in language very much like that above, appears in seventeen separate primary source documents coded for this dissertation, originating from nine different early partner institutions. Interestingly, however, all of these were public documents, and not interview accounts. Indeed, only one interviewee – a Googler – actually mentioned the word “mission” with respect to libraries.¹³¹ Still, despite the absence of the term within their utterances, I have little doubt from all that they *did* say (as well as from

¹³⁰ Where applicable. Not all of the Google partner libraries are part of an academic institution.

¹³¹ And in fact, many of the mission-related thematic codes utilized here – widespread access, public interest, sharing the wealth – were more frequently applied to official public-facing documents such as press releases and FAQ pages than they were to the depictions of motivation offered by the librarians I interviewed. In large part, of course, this reflects the much greater number of such documents in my research corpus, relative to interview transcripts. However, even accounting for this imbalance, the disparity is disproportionate. If I had to venture an explanation for it, I would note first that these mission-related rationales seem especially strategic when directed outward by university administrators and PR departments: given the controversy stirred up by the initiative, these direct rhetorical linkages to university and library missions – particularly given the lofty ideals those missions explicitly strive to advance – serve as strong rationales justifying the institution's participation to both local and external stakeholders. And second, on a more personal level, in my years working in libraries, I have observed an exceedingly high level of pragmatism among librarians. (The “sensible shoes” stereotype of librarians exists for a reason; if you are unfamiliar with this stereotype, a quick web search of “librarian sensible shoes” – no quotes – will show you what I'm referring to.) Many like nothing more than to make tangible, useful things happen in the information sphere, and they are often tasked with doing so for far less money than would be ideal. This is not to say, of course, that librarians are not also idealistic – many, if not most, certainly are; for more than a few, the profession's ideals are what make its relatively low pay and stodgy public image tolerable – but when directly asked about their institutions' motivations for participating in the Google Books Library Project, the librarians I interviewed overwhelmingly cited the pragmatic benefits to be gained, to a much greater extent than they discussed more idealistic goals.

my less formal interactions with many involved on the library end of the project) that the idealistic mission of libraries, to preserve and provide access to information, in service to present and future generations, formed a crucial backdrop to their personal and collective support of the project. The particular linkages between Google Books and the two pieces of this mission – access and preservation – are discussed in turn below.

2.2.1.1 *Broadening Access to Information*

Though broadening access to information may seem like a simple concept, as a motivation for library participation in Google Books it reveals a fair degree of complexity. In fact, different libraries and librarians described this increase in access as proceeding in quite disparate directions: while some waxed rhapsodic about their collections stretching out to reach users in remote corners of the globe, others focused more closely on the expanded forms of access that digitization would offer to their local user community – and some did both. In part, this is a question of imagined audience, which will be addressed in the next major section of this chapter; however, to the extent that increased access acted as a motivation, it can be divided into two basic pieces – global diffusion and local service enhancement – with an element of *noblesse oblige* bridging the gap between the two.

For some participating institutions, providing public – even global – access to information is an explicit part of their mission. The clearest example of this is Oxford’s Bodleian Library, for whom this element of mission extends back to the 17th century, when it was referred to as “the publique library of the University of Oxford” because it allowed in non-university users at a time when such liberality was extremely uncommon. And, even today, more than 60% of the Bodleian’s registered users have no direct affiliation with Oxford (Carr 2005b). As such, the Bodleian’s decision to pursue global public access to its collections via digitization can be seen as a direct extension of an extremely longstanding mission: as then-Bodleian Librarian Reg Carr expounded at the project’s outset, the library’s digitization partnership with Google “[fit] perfectly as a key modern element of what [had] been a strategic aim for the Bodleian for the whole of its existence: to bring its great collections to the wider world” (2005b).

And beyond the Bodleian, this type of argumentation is echoed in statements from several other library partners, including the University of Michigan (Coleman 2006), the New York Public Library (New York Public Library n.d.), Stanford (Stanford University Libraries 2006), the Bavarian State Library (Giersberg 2007), the University of Texas (Barnett and Willmann 2007), Cornell (“Cornell University Library Is Newest Partner” 2007), and Princeton (Princeton University Library 2010).¹³² For example, the Deputy Director General of the Bavarian State Library at the time of their partnership agreement, Klaus Ceynowa, explains that

¹³² Though perhaps with a few hundred years less historical depth in all of these cases.

As an international research library we don't have a mandate geared towards a specific user group. Rather, with our unique collection we have a commitment to research and teaching, to studying and science. In this respect, the Internet would seem the most natural medium for a traditional research library like ours. It is one of our prime objectives to make our holdings – provided it is permissible under copyright law – available digitally worldwide as soon as possible. (Giersberg 2007).

Though not all participating libraries hold public access as an explicit part of their mission, many do, and for those, as for Google, participation in this large-scale scanning project seemed a natural extension of that mission.

Further, the connection between Google Books and libraries' access mission was echoed in interviews with individuals at several institutions, in terms both general and personal. At the more general end, one librarian submitted: *"I think for the libraries, the dream of having the digital literature of their libraries available to as many people as possible was, um, obviously a motivation"* (P2), while another elaborated:

Broadly speaking, it is the foundational philosophy of librarianship. "Free the bound information." You know? Get what we have out to those who want and need it. The... so, the libraries at Harvard or Michigan or whoever else participate in this... I think, from the earliest stages of even imagining a Google project, were really on its side. As a way to get physical research material out to the world in ways that we could never hope to otherwise. You know? That scholar in South Africa, who needs a snippet from a medical dictionary.... Or you know, whatever. And as a notion about how information ought to be as freely accessible and available as possible, that's what libraries are all about (P13).

Among libraries engaged in scanning items still under copyright, a fine, yet significant distinction appears within the access motivation, between supporting discovery and enabling actual access to materials. As one Stanford librarian explained:

[One principal reason for Stanford's participation] was to get [the content] indexed, so that people could know that a book exists that has this information in it. And certainly yes we would love to have people be able to read it, but that was not... That was not a first motivation at all (P17).

While getting the institution's books *indexed* was a powerful motivation for participation at Stanford, providing actual full-text access to those books was felt to be agreeable, but decidedly secondary.

Still, whatever the finer distinctions here, the language in these passages – *"obviously," "certainly," "foundational," "what libraries are all about"* – begin to indicate how taken for granted it is among librarians that broadening public access to information is and should be among the institutional goals of libraries in general. While the parallel mission orientation of the Googlers involved likely makes them outliers among tech industry workers (at the very least, that orientation does not go without saying, and did not go without saying

within the interviews there), a commitment to mission is much more the rule among those who work in libraries; almost a condition of entry into the profession. Yet, as noted, it also remains personally felt, particularly with regard to this project. As one partner librarian explained, "*I just think, you know, come on, as a scholar, and a librarian, and a public service oriented guy, you know having information available is really [...] You know, the democratic impulse is just compelling, y'know, to me, personally*" (P3).

Still, as noted at the beginning of this section, libraries – and especially research libraries – are not uniform in their orientation toward public service. And in fact, the public statements of several Google partner libraries focus at least as much on improving the access available to their local user communities as on opening up generalized access to the world; in some cases more. Such local benefits were often cited directly alongside assertions of the value of broader public access (e.g., Harvard University Library 2004, Colvin 2006, University of Wisconsin 2006, Princeton University Library 2010, Biblioteca Complutense), appearing through parallel grammatical constructions and emphases to place equal weight on the interests of the libraries' local communities vis-à-vis the interests of the entire rest of the world. For example, both Stanford and the University of Michigan emphasize the project's value in creating a digital corpus for local use in the construction of novel discovery tools, the pursuit of new forms of textual and computational research, and the improvement of library services for affiliated students and faculty ("Stanford and GBS: Statement of Support" 2005, University of Michigan). Other library partners also seek to support their participation in the GBLP by citing the value of the program in general, and of Google's implementation of its Book Search tool in particular, to the local research and instructional missions of their institutions (e.g., Biblioteca Complutense, University of Virginia Library); in this respect, participation in the project is portrayed as a contribution to a public good that is already proving beneficial on the local level.

Bridging the gap between libraries' public and local foci with regard to access is a pervasive undercurrent of *noblesse oblige*; a sense within many partner libraries that the unique wonders of their collections ought to be shared with the world, and yet that such public sharing went beyond the call of duty, and proved goodwill and generosity of their leadership. After all, while for some participating institutions, such as the Bodleian Library and New York Public, service to the public forms a central part of their mission, this is not true of many others. The explicit mission of most academic libraries is to serve the needs of their university community, and not the surrounding public. This is especially so at private institutions, and within the data gathered for this case, is nowhere more clearly demonstrated than at Harvard. In fact, the issue of *noblesse oblige* as such was first pointed out to me in an interview with a Harvard librarian, who noted that,

the ethic in a sense at the higher levels of library thinking at Harvard have been that, well, "you can come to us and we'll do what we can to assist in a sort of noblesse oblige

way. But don't expect us to..." I don't know... "do it the way you think we should" (P13).

And she goes on from here to explain how Harvard's Open Collections project – a local digitization effort – helped to pave the way not only for the university's participation in Google Books, but also for its ambivalent attitude toward the project:

And I think that [project] was what made it possible to envision going into the Google project. The complication, though, is that the Open Collections project was a Harvard project. All about special stuff. That Harvard was doing you the favor of making available. You know, our gift to you. Don't count on anything else, but, you know, use this in good health. Whereas the Google project was a much more, call it democratic, approach to the way that library resources should be commonly and customarily available no matter what (P13).

Moreover, this informant's perceptions of the somewhat patronizing atmosphere at Harvard are supported by other statements from the university. For example, librarian Sid Verba's announcement to the Harvard faculty, the day before the GBLP was officially announced,¹³³ begins by declaring:

As all of us know, Harvard's is the world's preeminent university library. Its holdings of over 15 million volumes are the result of nearly four centuries of thoughtful and comprehensive collecting. While those holdings are of primary importance to Harvard students and faculty, we have, for several years, been considering ways to make the collections more useful and accessible to scholars around the world. Now we are about to begin a project that can further that global goal – and, at the same time, can greatly enhance access to Harvard's vast library resources for our students and faculty (2004).

While no other institution goes quite so far into superlatives to describe the wonders of its collection and its own generosity in sharing it, a thread of self-congratulation for sharing the wealth of their collections beyond their institutional boundaries nonetheless appears throughout partners' public statements ("Stanford and GBS: Statement of Support" 2005, Coleman 2006, Cliatt 2007, Ghent University Library 2009, University of Wisconsin Libraries). Yes, they seem to say, we believe that public information access is wonderful – but aren't *we* also wonderful for sharing our treasures with everyone?

2.2.1.2 Preservation

Preservation also forms a huge and pervasive theme among libraries' depictions of their motivations for participation in the GBLP; 43 excerpts in documents from twelve institutions¹³⁴ explicitly contain some form of the word "preservation,"¹³⁵ and many place

¹³³ The timing of which, as noted, was in itself revealing.

¹³⁴ Specifically, Bavarian State, California, Columbia, CIC, Ghent, Madrid, Michigan, Oxford, Stanford, Texas, Virginia, and Wisconsin (P2, P17, University of Michigan 2004a, "Stanford and GBS: Statement of Support" 2005, University of Michigan 2005, Coleman 2006, Colvin 2006, Barnett and Willmann 2007, Columbia University Libraries

its importance above all other factors, including access. As a 2005 Stanford public statement avers,

PRESERVATION is one of the vital social roles that libraries, particularly research libraries such as Stanford's, provide. Stanford's primary intent in obtaining a digital copy is to ensure the preservation of our library's resources. Library collections are vulnerable to catastrophic losses from fire, flood, or earthquake. For example, Stanford libraries suffered flooding in 1978 and 1998, resulting in damage or loss to many thousands of books. These losses were minor compared to the more recent and tragic flooding of the University of Hawaii Library and the public and university libraries of New Orleans....A digital copy of works ensures the future of not only Stanford Libraries resources, but of the collected works of our society, our civilization ("Stanford and GBS: Statement of Support" 2005).

Notably, however, it is also clear that the preservation focus within this project was much more on preserving the *information in* the books rather than preserving the books themselves, as artifacts. As one librarian recalled, it became clear early on that "*the level of digitization was not up to the prevailing standard of libraries,*" and yet, he observed that through this effort,

libraries began to be aware [...] for the first time that sustaining that level of preservation activity at scale was probably not possible, and some kind of compromise realistically needed to be effected. But nonetheless it was an initial stab at creating a fallback digital copy for content that otherwise would not have had a digital surrogate (P2).

Particularly for institutions in California, which this same participant quipped "*were all pretty much on the same faultline,*" a significant part of the appeal of Google Books was that it would provide the libraries with the opportunity to make the informational content of their collections redundant, and thus redundantly storable in any number of less hazard-prone locations (P2, P17).

Still, one might debate the extent to which "true" preservation can be effected through digitization (which tends to produce quite ephemeral formats), or, even moreso, the wisdom of going about that digitization in a low-cost, high-throughput manner (which tends to increase the number of errors in at least the first-run output) – and indeed, many have argued this latter point in particular (e.g., Vaidhyanathan 2005, Coyle 2006, Duguid 2007, Townsend 2007, Nunberg 2009b). Some of the quality issues with the project's scans will be discussed in section 4.2.3.1, below; for now I will simply note that, given a

2007a, Columbia University Libraries 2007b, Giersberg 2007, "Settlement FAQ" 2008, Committee on Institutional Cooperation 2008, Cunningham 2008, "Stanford's Role in Google Books" 2010, Keller 2011, University of Michigan Libraries 2011, Biblioteca Complutense, Committee on Institutional Cooperation, University of Michigan, University of Texas Libraries, University of Virginia Library, University of Virginia Library, University of Wisconsin Libraries, University of Wisconsin Libraries)

¹³⁵ E.g., preserving, preserve, preservación.

sustainable plan for data and format refreshing, even this kind of relatively low-resolution digitization can provide a significant degree of insulation from the kind of total loss of information resources – both content and container – that occurred in New Orleans during Hurricane Katrina (whose destruction of libraries several participants cite directly; e.g., Coleman 2006, Colvin 2006, "Settlement FAQ" 2008), or that could well occur in the event that “the Big One” finally hits the San Andreas fault. And of course, even setting aside the risk of major disasters, making a digital version of book content available can also aid in preservation simply by reducing the usage load on the original; for many, the digital avatar will be just as useful, or better, even if it is not a 100% faithful reproduction of the original (e.g., Carr 2005c).¹³⁶

2.2.2 Pragmatic Aims

As noted above, the pragmatic motivations for participation can also be grouped roughly into two groups: one focused internally, on enhancing and expanding existing library functions, and one focused externally, on bolstering institutional reputations and ongoing library relevance in the digital world.

2.2.2.1 *Support and Amplify Existing Library Undertakings*

Libraries perceived many ways in which participating in Google Books might help to advance existing institutional priorities. Most directly, the project presented a clear opportunity to vastly expand the scale and scope of existing library digitization initiatives, well beyond what the libraries felt they would have been able to accomplish on their own. The first University of Michigan FAQ about the project asserts that even given the demonstrated strength of the UM Libraries’ existing digitization programs,

we [at UM] know that only through partnerships of this sort can conversion of this scale be achieved. Our program is strong, and we have been able to digitize approximately 5,000 volumes/year; nevertheless, at this rate, it would take us more than a thousand years to digitize our entire collection (University of Michigan 2005).

An interviewee at Stanford concurs with this assessment, explaining that the initiative was felt to be equivalent to “*local digitization projects on megadoses of steroids*,” and that the sense was that Stanford could “*make happen through Google what we would take multiple generations to do in the ways we were doing it up to then*” (P17). Similar rationales, extolling the benefits that had been reaped through prior digitization, and averring that this partnership presented a rare opportunity to massively expand those benefits, also appear in statements

¹³⁶ Indeed, I repeatedly found this to be the case myself while doing research for this dissertation, especially on the historical cases. Many of the resources I needed were available to me in paper locally, but given a free digital PDF through Google Books or the Internet Archive, even with a few stray scanner fingers or blurred pages, I rarely found the need to consult those physical editions directly.

or interview accounts from the University of Virginia, Columbia, the University of Wisconsin, and the CIC (P9, University of Virginia Library 2006, Columbia University Libraries 2007a, University of Wisconsin Libraries).

Perhaps less obvious and more intriguing, however, are the myriad pragmatic, yet *indirect* benefits that libraries and librarians also ascribed to the project; that is, ways in which their participation advanced internal priorities beyond digitization itself, including collection and stacks management, interlibrary loan, and cataloging. The participating institutions varied widely in which of these priorities they emphasized. Collection and stacks management, for example, was highlighted most clearly in materials from Michigan, California, and Oxford (P3, P8, P10, P16, Carr 2005c, University of Michigan 2005). A statement from one leader in the University of California's Google partnership is representative: he avers that in his mind, mass digitization

was about access, was about preservation, but it was also, at least in my personal view was really very extensively about collection management. Because, guess what, the real challenge to libraries, that was pretty obvious then, it's even more now, is they're gonna run out of space, they're gonna run out of money, and if they haven't solved the problems of legacy collections, they're going to have a real struggle going forward (P3).

The Bodleian's Reg Carr also picks up on the idea of reducing costs by reorienting usage of the library's physical shelf space: as he notes, given mass digitization,

it may well become possible, as we develop shared retention policies with certain UK national partner-libraries, ...to save some physical storage costs for some of the more common materials (and especially perhaps as the archiving of digital materials becomes more robust and reliable for the long-term) (Carr 2005c).

Making books available online is rarely, if ever, mentioned as a rationale for fully deaccessioning or discarding materials (indeed, that idea is specifically rejected in both the Carr piece and the Michigan FAQ cited above). However, shelf space in libraries' open stack areas is both limited and valuable, and significant savings can be achieved by moving lesser-used materials into remote storage (e.g., Courant and Nielsen 2010). The type of relatively comprehensive digital, public, full-text availability of public domain materials that the Google Books Library Project made possible makes it much easier for libraries to justify shifting their physical copies of these materials offsite,¹³⁷ to make room for new collections or new library functions.¹³⁸

¹³⁷ It's worth recalling at this point that most books in the public domain are quite old, and that due to their age, they are likely to be among the lesser-used items in the collection. In the U.S., books published prior to 1923 are generally in the public domain, as are any that were published between 1923 and 1963 without complying with formalities (notice, registration, and renewal). The Copyright Review Management System, based at the University of Michigan, has been used to review more than 200,000 works from this period, and has been able to establish that 54% of them are in the public domain and can be released in digital form through HathiTrust (Adler and Levine 2013).

¹³⁸ As I have argued elsewhere (Jones 2013a).

Other indirect pragmatic motivations cited by participating institutions seem to have been more idiosyncratic. Oxford, for example, is alone within the dataset in expressing an intention to use the products of the Google scanning to reduce the use of physical collections for interlibrary loan (Carr 2005c). And while the Universidad Complutense de Madrid anticipated from the outset that the project would be a boon to local cataloging efforts (Biblioteca Complutense), for other institutions this benefit was less an initial motivation than a unanticipated bonus. For example, interviewees at Harvard and Oxford noted that in retrospect, their institutions realized significant gains in cataloging and overall understanding of their collections due to their participation (P13, P16). At Oxford, moreover, this boon to collection organization included the discovery of thousands of books the library didn't even know it had. As my informant there explains,

We were able to catalog a lot of books that had never been cataloged before because you know books that were, you know, for some reason when they came into the library in the 19th century were never cataloged and sat on shelves almost hidden. And so we were able to... because we were handling so many 19th century books, we were able to find, you know, several thousand of these, almost by accident, and cataloged them and provided access to them (P16).

And finally, an informant at Stanford emphasizes the extent to which participation in Google Books was seen as a likely catalyst for moving forward other, local digital initiatives – especially the development of the Stanford Digital Repository (<http://lib.stanford.edu/sdr>) – and suggests that the library's participation in the GBLP was crucial to securing the necessary funding for that project, commenting, “I'm sure the funding that we got for [the SDR] was very directly related to the notion that we would be having all these books” (P17). And although this type of local digital project is not directly cited as a motivation by any other informants, it seems clear to me that many, if not most participating institutions had some plan to utilize their new, Google-scanned digital collections as a springboard for extending local digital initiatives – if only by adding links to digital versions of locally-held books to their online catalogs (e.g., New York Public Library n.d.).

Large-scale digitization, in sum, was seen within participating libraries as providing a diverse range of pragmatic, local benefits, via both the digitization itself and the various procedural improvements that either needed to occur to make the project happen or were foreseen as outgrowths of possessing thousands – or millions – of digital books at the library level.

2.2.2.1.1 Cost Coverage

Arguably, of course, similar benefits could have been gained through any large-scale digitization initiative – had any others of similar scale existed. But there was a reason no others did exist, and although there are many reasons for that, I would argue that a primary

one was money. This kind of project inherently requires the participation of many institutions, as well as the aggregate expenditure of at least several hundred million dollars.¹³⁹ And within the library community, before Google's project existed as a proof of concept, it was exceedingly difficult to build consensus around funding such ambitious collaborative digitization projects. One interviewee provides a useful anecdote to illustrate this point:

in 2004, in Arizona, while the [Google] project was still under wraps, the ARL directors got together to consider a business plan that had been assembled to digitize U.S. federal government documents, and they rejected it emphatically. They rejected it for a whole lot of reasons. Um, it was, it was infeasible. It was not our responsibility. It was problematic with regard to fidelity in the public record. They didn't [...] they didn't do it. They didn't want to do it, and they said no. [...] They didn't come to it lightly. But that was the mood of the community at the time. "No." "No" for all sorts of reasons. "No." And that held us back (P8).

This interviewee goes on to explain that this mood of "no" prevented libraries from taking the collective action that some GBS critics argue should have occurred without Google's involvement (e.g., Vaidhyanathan 2005). And that is likely so. But there was also a chicken and egg problem. Nobody in the library community would get together to fund truly mass-scale digitization, because nobody had proved that it was useful or practical to do so – but the only way to provide that proof was to do it, which cost money. And Google, as noted, brought its wallet.

Several public statements from libraries point to the significance of Google's proffered cost coverage in swaying institutional decisions to participate, with many averring that their institutions could never have afforded to pursue digitization at this scale and pace under any other circumstances (e.g., University of Michigan 2004b, Coleman 2006, Giersberg 2007, Committee on Institutional Cooperation 2008, Wittenborg 2009). And interview accounts concur. For example, at the University of California, where digitization was already underway with the OCA when Google stepped in, one informant explains that financial considerations played a substantial role in the University's decision to work with Google also, in parallel with the OCA:

[UC] recognized on one hand the tremendous advantage that Google brought to the table. That, um, that they brought funds, they brought support for scanning. [...] Obviously, alternatively, Yahoo! and Microsoft came to the table... but with relatively limited amounts of money. Even...the monies pledged by Microsoft were sort of [...] relatively trivial, obviously, compared to the amount of funds that Google implicitly had put on the table, uh, by pledging to entertain digitization of entire collections. And, and so... from the perspective of senior management at a university, um... to, to cease entertaining the possibility that Google would scan an entire university library collection would have been detrimental to the interests of the university, arguably (P2).

¹³⁹ See cost calculations in footnote 125.

The cost calculus also proved persuasive within the CIC, in the context of intra-consortial negotiations prior to the agreement. As a leader within that consortium recalls,

We had this conversation in 2006, 2007, when we were doing our CIC agreement with Google, and one of the Provosts was actually quite opposed to the whole project. And, you know, on commercial grounds. I woulda called it naïve myself, you know, but just... you know, "we're the university of Blah! We're not going to sign Google's agreement and do their work for them! What's in it for us? And what would it cost us to do that anyway?" And you go, "Okay. Let's figure. Let's figure this, let's figure 60 dollars a book... let's figure..." uh, you know, and then there's like dead silence at the other end of the line. They go oh, well, [laughs]. That's just never going to happen. We would achieve a... you know... we would be... it'll be cobbled-together resources at that level, to do what Google has done (P9).

And of course, the CIC ultimately did sign on, with some of the strongest early skeptics among its membership eventually evolving into equally passionate project advocates (P9). Perhaps the bluntest statement of the motivational power of cost coverage, however, comes from the Bodleian's Carr, who asserts quite frankly that "One of the major benefits to us in Oxford, of course, is that Google is willing to pick up virtually all of the costs of the Oxford operation; and this, if you like, is the essential quid pro quo for us that makes it such a compelling opportunity" (2005c).

Before Google stepped in, it was difficult for libraries to rally internal or collaborative support for this type of large-scale project, especially considering the enormous expense of doing so; by offering to pay, Google brushed past that sizable hurdle, and made what had previously seemed infeasible seem not only doable, but attainable within a decade. As the President of the University of Michigan declared in a 2006 speech to the AAP,¹⁴⁰ Google's venture made "the global library...no longer a question of 'whether,' but rather 'how' and 'when'" (Coleman). Though I sincerely doubt that anyone at Google had such a goal in mind at the outset, by bankrolling the Google Books Library Project, the company not only motivated library participation in its project, but also catalyzed a fundamental shift in the way that both libraries and the public think about large-scale digital access to books.

2.2.2.2 Improve Library Reputation and Ongoing Relevance

The final piece of the mosaic of library motivations relates to the reputation and relevance of libraries, both individually and collectively, in the digital world.

On a local level, many universities and libraries saw collaboration with Google as a way to improve or gain competitive advantage and reputation vis-à-vis peer institutions. For many

¹⁴⁰ The AAP, it should be noted, had filed their lawsuit against Google just a few short months earlier, in October 2005. Delivered in February 2006, that speech was a gutsy move. (Though, in the interest of full disclosure, I should also note: I worked on the research and speaker prep for that speech, as a Master's student working in UM Media Relations and Public Affairs.)

institutions, being invited to participate in GBLP felt like a significant recognition of the importance of their collections; an acknowledgement that their library was a leader and a center of excellence within the field. As one participant put it, *"You know Google signed up the prominent libraries around the world. [...] I think people were flattered to be included, for the most part. It sort of said something about the level of your collections"* (P18). And another mused, *"I think we wanted to be a leader in the libraries in the UK. That's what we wanted to... you know, we've always felt that we are a leading institution in our industry and we wanted to keep up with that"* (P16). And several other interviews and public statements, concur, citing the importance of institutional competitiveness and leadership as factors in their decision to sign on with Google – in many cases quite directly (e.g., P2, P9, Coleman 2006, University of Wisconsin 2006, Keio University 2007, University of Virginia Library).

More collectively, there's a strong sense throughout the data that libraries saw partnering with Google as an opportunity to modernize their collections and their institutions; to keep their resources and services relevant to their users (and especially students and other young users) in the digital era. It bears recalling at this point that Google was "cooler" in 2004-2005 than it is now. They had only just gone public, in one of the most successful IPOs ever (and with one of the quirkiest IPO statements ever (Page and Brin 2004)). Most of their more controversial decisions and products still lay ahead of them (e.g., moving into China, Google Buzz, Google Glass...). Nobody was yet calling them "the new Microsoft." All this is to say that, at the time, Google still retained a great deal of their shine as the "not-evil" player in the tech industry. And libraries, well, people had started spouting off about their irrelevance and predicting their imminent demise decades earlier, and that trend had only grown stronger with the explosion of the public Web in the 1990s, and especially the entry of Google in 1998. As such, it is perhaps not surprising to see many libraries and librarians declaring that they had signed on to the GBLP to maintain their relevance in the digital world – how better to accomplish this, after all, than by partnering with the coolest and most relevant kid in Silicon Valley? The Bodleian's Carr links his institution's participation in this project especially strongly to the need to stay relevant in the digital world:

With information in digital form bringing such huge changes into the world of human communications, the continuing value of the Bodleian will be judged by the extent to which it builds on its unique heritage by developing a completely new paradigm of electronic access to all the knowledge that it contains in physical form. And this is one of the reasons why 'digitisation' is now as much a part of library life in Oxford as it is anywhere else in the developed world. ... [We] see the mass-digitisation of our printed collections as one important part of the long, and evolutionary, process of updating Gutenberg's legacy by the application of the latest information technology. Digitisation, in our view, is not replacing Gutenberg so much as modernising it (Carr 2005a).

This argument is reinforced by the University of Michigan's Coleman. She recounts a Ford executive's exhortation to "Change or die" within the rapidly changing auto industry, and applies this maxim to universities, given the transformations being wrought by the Internet:

New technology is disrupting all segments of our society. Newspapers and TV networks are trying to figure out how to make money with online editions. Hollywood is experimenting with simultaneously releasing movies to theatres, DVD and cable. Cell phones are ubiquitous. For better or worse, they are shaping how, when, and where we communicate. Universities are not islands in this sea of technology. We must change with our students, and that means embracing the Internet and all it can, and does, offer (2006).

And, as she goes on to explain, the University of Michigan considers digitization in general, and digitization with Google in particular, to be crucial to the University's ability to continue to thrive and lead in the digital environment – and that assessment is echoed throughout both UM's public statements on the project (e.g., University of Michigan 2004b, University of Michigan), and public statements from other partner institutions (e.g., "Stanford and GBS: Statement of Support" 2005, Colvin 2006, Keio University 2007, Committee on Institutional Cooperation 2008, Cornell University Library, University of Wisconsin Libraries). As more and more information seekers take their questions to the open web, it is necessary for libraries – and their materials – to remain in those users' line of sight. Partnering with Google Books was viewed by many librarians as one way of "going where the users are," and thereby perpetuating the relevance of their institutions in the digital world.

3. Definitions

The section above describes the diverse motivations expressed by Google and its early library partners for initiating and engaging in the Google Books Library Project, largely tied to their respective institutional missions and internal strategic and pragmatic objectives. As in previous chapters, this section will explore some of the ways in which these same stakeholders envisioned their project's potential user base and collection scope (in that order), and how those visions relate to the underlying motivations described above.

3.1 Intended Users

Unlike the historical chapters, those dealing with the digitization cases will not explicitly assess the numerical size of the projects' user base, for two reasons. First, and most significantly, there is simply no reliable quantitative data upon which to base any such assessment in either case. Google only began releasing usage numbers *to its library partners* around the time I was executing my interviews in the fall of 2011, and even then, only in a limited way, and subject to non-disclosure agreements that the librarians felt prevented them from sharing those numbers with a researcher (P9). And second, the scope of the user

base for Google Books – as for the Open Content Alliance – was much more explicitly conceived as “everyone” – as many people as possible, worldwide, with no consistent requirements for things like user registration or set boundaries on geographic service areas – than the historical cases ever were. As such, the most I would be willing to venture with regard to the intended and/or potential size of the Google Books user base is that it is basically coterminous with the number of people who use Google.¹⁴¹ And although available estimates of the size of that user base vary, it seems clear that the number of searches executed on the site *per day* is in the billions.¹⁴² Although clearly only a fraction of these searchers will find their way over to Google Books, their presence on the site leaves open the potential for them to do so.

Further, to a certain extent, this uncertainty regarding size can also be extended to the *composition* of the GBLP user base. It was quite common in the interviews for participants, especially at partner libraries, to have to take some time to think about their response to the question about who they envisioned using the service. In part, this was because many – though not all – had simply never previously considered this question in any detail. So, for example, one starts his response with “*I don’t know who I would picture*” (P3) and another admits, “*I haven’t thought about it a lot*” (P9). In addition, as with the question of the size of the user base, there was also some data-gap-related uncertainty about who was actually using the service, at least for the librarians. Since Google was the only one with direct access to actual usage data, and they had only recently begun to share it at the time of the interviews, library partners had some difficulty gaining an understanding of what sorts of internet users were using which pieces of their collections and how, even for the purposes of championing the project within and beyond their institutions. For example, one CIC librarian laments,

Well, this is very sad. Because we know almost nothing about users. For this. And have struggled forever to, you know, even come up with anecdotes or things that we could relay if we were doing a road show or out there talking about the project, or why we thought the project was good (P9).

Although at the time we spoke, Google had begun to share some data (after considerable lobbying by its library partners), the uncertainty wrought by years of lacking that data nonetheless remained for many of the librarians I spoke with.

¹⁴¹ As, indeed, some interview participants explicitly suggested to me (P12, P13, P18).

¹⁴² Google itself says that it “answers more than one billion questions from people around the globe in 181 countries and 146 languages” on a daily basis (Google n.d.-a), but does not provide a date for this number. And that date may well be quite relevant: quantitative data aggregator Statistic Brain places the figure for average daily Google searches at over 5 billion for 2012, up from about 4.7 billion in 2011, and 3.6 billion in 2010 (<http://www.statisticbrain.com/google-searches/>). If these statistics are accurate, it would lead one to believe that the Google figure cited above is either three or four years out of date, or reflective of a very conservative mentality toward data reporting (and the latter is quite plausible, as will be seen in the discussion of collection size below).

Still, despite these ambiguities and uncertainties, the data reveal several compelling themes with regard to the intended composition of the GBLP user base. In the sections that follow, I will discuss three of these themes: (1) the Global Public, (2) Favored Users (the Consumer, the Scholar, and the Genealogist), and (3) Not-So-Favored Users (the Non-Connected, and after the proposed settlement, the Non-American).

3.1.1 The Global Public

The desire to truly, literally, serve everyone and anyone, worldwide, is most frequently attributed to Google rather than its library partners, though the theme also appears frequently in public statements from partner libraries. Although the librarians I spoke with also appreciated (and even delighted in) the project's global reach, they were far more likely to specifically focus on scholarly uses and users as their dominant target group (or if not dominant, then at least equivalent to everyone else in the general public in importance). This difference in focus was explicitly noted by one Googler I interviewed, who observes that while Google strives "*to serve the general user around the world...libraries often have a more academic or scholarly audience,*" and goes on to elaborate that:

Google in general tries to appeal to the widest, broadest um segment of users as possible. And...my observation is that we tend more to try to build, you know... tools like APIs, and things like that, that allow subsets of users to come and craft things the way that they want, without trying to specifically design a particular site or a particular product for a particular class of user. More like give users the tools. Um, I, you know, but I would say that if you compare Google Books to HathiTrust for example, you know there are design choices that we've made that they haven't made and vice-versa. That probably, um... reflect the idea of what a user values. And so. You know, maybe they have a different idea of a user than we do (P12).

For Google, it seems, the concept was much less about striving to serve any specific user type or demographic than it was about trying to make the product usable and extensible by as many different sorts of users as possible. And indeed, the observations above were echoed in several other interviews, both at Google and at libraries (P4, P8, P10, P17, P18). For example, a strong echo of the above depiction of Google's motivations resounds in this statement from a librarian at the NYPL:

Google was looking at the general public, right from day one. So I don't think that they in their mind necessarily looked at this sort of challenge ahead of them and said "oh there's stuff here that maybe 10 people in the world would want to see." I think they just looked at it and said "these are things that everyone would want to see," or by mass numbers, you know, millions of people would want to see, hundreds of millions of works, and maybe one work will be viewed hundreds of thousands of times, and one never. But. That's not for them to tell. They have a very sort of mass-market, mass-process view of things. Just put out there, so it can be used on any device, by any person, anywhere, and let them figure out what they want to do with it (P18).

This passage reinforces the perception that Google's focus was always on the whole world, as well as the sense, alluded to in the discussion of motivations, that there was a certain innocence and optimism about the Google staff involved in this project – something to the effect of, “Of course the whole world will want to see these old books! Who wouldn't? Let's just throw it all out there for everyone to play with, and see what they do!”

Still, as noted, the project's global user base also features prominently in many of the libraries' and universities' public statements about the project – and not just as Google's vision, but also as their own. The UM's Coleman, for example, is quoted in the original announcement of the project as saying, “We believe passionately that such universal access to the world's printed treasures is mission-critical for today's great public university” (Google 2004). And several other partners laud the project's capacity to throw their collections “open to the world” (Armstrong 2011), “provide global access to our library resources” (Marsh 2007, quoting the President of Cornell University), and facilitate “scholarly research on a world-wide basis” (University of Texas Libraries n.d.). However, it is also worth noting that in many of these public statements, the global public is mentioned in the same breath with scholars, but explicitly separate from and on a par with them – as in, “scholars and the general public” (Cornell University Library) or “the academic community and far beyond it” (University of Michigan 2004a). Such parallel constructions, placing the community of scholars on a level with everyone else in the world put together, hint at the libraries' central vision of the user as scholar, which will be discussed further in the next section. The project's global reach may have been significant and inspiring to librarians, but the core mission of research libraries has historically been to serve scholars, and that emphasis was reflected in their statements regarding Google Books.

3.1.2 Favored Users

Delving more into specifics, the data reveal three types of user as being especially favored within or benefited by this project: the consumer, the scholar, and the genealogist. I will discuss each of these briefly, in turn.

3.1.2.1 *The Consumer*

As may already be clear from the foregoing discussion of the global user, on the Google side, Google Books users are mainly conceived as a segment of the overall consumer search market, to which the company doesn't specifically ascribe any particular educational or scholarly intent. This is the type of user that one Googler interviewed describes as “average Joe searching the web:” the type of user who might

enter a query on the web, and one of the results happens to be a book, for which they can either read the whole book, or they can do a preview (for these people I think in general snippet-view books are, um, less relevant, because of the limited access). So, one of the examples I've always enjoyed, was one of the queries we had was, um, “concrete fountain

molds.” And, they found a public domain book, on concrete fountain molds, that they then read, and spent quite some time and downloaded, and y’know, with a 1906 book, and...a statement I’ve made many times is I suspect concrete fountain molds have not changed radically since 1906. ... And so as such [...] that was a good result (P4).¹⁴³

As this participant also noted, this vision of the user became especially prevalent after Google began what it calls “blending” for Google Books: that is, integrating hits from books.google.com into general web searches run on its main website, when such hits seemed relevant – as in the case of the “concrete fountain molds” query described above. Several librarians also emphasized Google’s consumer focus, over and above any particular interest on their part in more “serious” scholarly uses (P2, P9, P13, P17). A comment from a Harvard librarian is thematically representative of these excerpts, though her assessment is more detailed than most:

I can think of an awful lot of recreational or consumer purposes to use Google, Google Books, Google Scholar for... and whatever you know the business model is, that might be ideal. But, it is hard for me to as a... probably, over-structured librarian, to see the Google projects as really useful tools for definitive research, teaching, or learning. ... I would suspect that they are thinking about users almost exclusively as one kind of consumer or another. Whether it’s a direct consumer who will click on “Get It At Amazon” if all they can see is a snippet... or who is excavating, say, prices of collectible automobiles or things of that sort. You know, a mode for facilitating certain kinds of economic behaviors through the use of Google broadly speaking as a knowledge management tool. Beyond that, it’s hard to understand, or hard for me to understand Google as a really serious tool for academic research. Even despite how much Google loves the fact that data mining is possible from the vast tracts of texts that they’ve captured (P13).

To a significant extent, it seems clear that although the Googlers involved had a very general sort of user in mind, their image of that general user was essentially consumer-centric, and particularly focused on the type of consumer Google was already used to dealing with in its search business. Both Googlers I interviewed expressed high levels of excitement when talking about how this project could open the information contained in books to Google’s existing search consumer base – not for any specific educational or scholarly purpose, mind, but simply because that information might be useful to them, and helping people find information that’s useful to them is foundational to both Google’s mission and its business model.¹⁴⁴

¹⁴³ Data note: several “uhs,” “ums” and “y’knows” were removed from this quote to improve readability. In my judgment they did not substantially affect meaning in any way.

¹⁴⁴ It does bear mentioning that this image of the user-as-consumer would undoubtedly have been much more highly emphasized had I also been asking about the Google Books Publisher/Partner Program, rather than limiting my inquiries to the Library Project; the interviews where that aspect of the project did come up indicate as much, as do Google’s promotional materials for that project (e.g., P12, P17, Google n.d.-c). Within that portion of the Books

3.1.2.2 *The Scholar (Local and Global)*

The finding that many of those involved with this project, especially on the library side, were particularly focused on scholarly and educational uses should strike no one as especially surprising. After all, the libraries that partnered with Google, especially in the early years, were *research* libraries: libraries whose mission of access and preservation revolves around serving the needs of scholarly communities (at varying degrees of breadth) in the present and far into the future. Still, there are a few distinct threads to tease out within this scholarly conception of the user for Google Books.

On the one hand, as noted, serving scholarly constituencies is a clear extension of the libraries' respective missions: most exist primarily to serve a local community of scholars, but many also have a strong interest in advancing research and teaching more broadly. These two elements come through quite clearly throughout the documentation of the project emanating from libraries. For example, the initial announcement of the project to Harvard's faculty explains:

While those holdings are of primary importance to Harvard students and faculty, we have, for several years, been considering ways to make the collections more useful and accessible to scholars around the world. Now we are about to begin a project that can further that global goal – and, at the same time, can greatly enhance access to Harvard's vast library resources for our students and faculty (Verba 2004).

Extremely similar sentiments appear in statements from the CIC, Columbia, Cornell, Keio, Michigan, Oxford, Princeton, and Wisconsin.¹⁴⁵ As one Googler summarized, where Google has a fairly general user in mind, "*libraries often have a more academic or scholarly audience.*"

project, Google and its partners are more in sync in viewing readers as consumers – though Google likely views them mainly as a consumers of search and advertising, while the publishers view them mainly as consumers of books.

¹⁴⁵ For example:

- "The digitization project will provide teachers, students, scholars, and readers around the world with an unprecedented ability to search, locate, and read books from the University's collections. The digital collection resulting from this project significantly advances Columbia's ability to serve its academic community, as well as readers worldwide." (Columbia University Libraries 2007b)
- "CIC Universities wish to digitize portions of their collections in a manner consistent with copyright law to facilitate use of their collections by scholars and the general public for education, research and related purposes" (Committee on Institutional Cooperation and Google)
- "As a result, materials from the library's exceptional collections will be easily accessible to students, scholars and people worldwide, supporting the library's long-standing commitment to make its collections broadly available." ("Cornell University Library Is Newest Partner" 2007)
- "The project will enable Keio University Media Center to provide intellectual assets of the recent era of Japan in an international basis, and to contribute to academic research activities in and out of the country" (Keio University 2007)
- "The online medium also allows the University to share its collection with researchers worldwide, a step very much in keeping with the University's unofficial motto of 'Princeton in the nation's service and in the service of all nations'" (Princeton University Library 2010).

And he goes on, "*I think the HathiTrust is a good example*" of the way in which libraries "*are taking these volumes that we're digitizing from their collections and they're making active use of them,*" in a mode more directly oriented toward academic uses (P12). And while there was also some equivocation among interviewed librarians about the actual effectiveness of the products of Google's scanning initiative for scholarly use (e.g., P13), there is also a fair degree of optimism about libraries' capacity to build better tools for scholars on top of the digital substrate produced by the Google digitization (e.g., P3, P8).

However, there is also a subtler way in which visions of a scholarly user base come through: namely, in ascribing basically educational intentions or purposes to users in the general public as well; in the excitement expressed at the thought that the general public would be able to use this great resource to *learn* and to *increase knowledge* (e.g., "Stanford and GBS: Statement of Support" 2005, Carr 2005b, University of Michigan 2005, University of Wisconsin 2006, Ghent University Library 2007). In this, the rhetoric surrounding Google Books is very, very similar to that of the historic public library initiatives already discussed. There is a strong interest in serving the public, but this interest is most keenly felt toward *a specific kind of public* – the kind that *wants to learn*. And of course, this makes a great deal of sense: the collections of major research libraries tend to contain far more cultural, historical, and educational materials than popular or entertaining ones. As such, simply by virtue of the collections being scanned, the segment of the public that will be best served will be the segment that is most interested in those types of materials. As expressed by one UM librarian, the sense is that by catering to academic users, this project will end up serving the public as well:

we serve academic users, and we found that over time in serving academic users, we serve the broad public. [...] [W]hen you do something on scale, you, you have more serendipity, serendipity of discovery and serendipity of use. So that famous beekeeping manual [in the Making of America Project]. [...] You know, it was just a pick, a pick from the shelf. It was something we owned and it got online, and it turns out to be a bible for those people. And... "for those people" – I'm sorry! For that... those people are some of our people! Right? One of our developers at the time was a beekeeper and it turned out to be very important to him. But you just run into these things randomly. It turned out one of the things we digitized in Making of America was big in National Treasure when the movie came out. It was significant in, uh, as a piece of, you know, piece of culture that factored into the movie. And then it became immensely popular. You can't predict those sorts of things. And so, it is, I think, my sense and my hope that in doing something for the broad population of academic users, we serve people at large. We make ordinary uses for non-academic purposes possible. ... We don't want to throw up walls, um, to say "Not this type of user," but to assume that in doing well for one community, we're also going to serve other communities. So, so the focus is primarily on our constituencies, with the recognition that in doing it well, we're going to serve other constituencies, too (P8).

The items that can be made available digitally in full text under current copyright law are inherently somewhat outside the popular mainstream – generally very old, fairly esoteric, and appealing only to a tiny fraction of the population – and tend to be much more effective for educational use than for pure entertainment. Recognizing this fact, it is perhaps unsurprising that those involved in GBLP should highlight the public educational capacities of the corpus, rather than its pure popular appeal.

Interestingly, to some extent, the concept of serving the public by first serving academics could be seen as a direct inversion of Ticknor’s early plan for the Boston Public Library, wherein the initial focus was placed on serving *less* academic, more general users, with scholarly constituencies only catered to after more popular needs were met. And yet, an explanation for this divergence springs readily to mind, relating to the prevailing state of information access in the 1850s vs. the early 2000s. That is, in Ticknor’s Boston, the major information problem was *scarcity*: there just weren’t all that many sources of information available to the public at that time, even in a place of relative information wealth like Boston. And because that scarcity was especially acute for those without access to institutions like the Athenaeum or the local universities’ collections, it made sense for the BPL to tilt the balance toward popular materials first, in order to serve the greatest perceived need. By 2004, however, in the context of the World Wide Web, this balance had shifted. The problem by then was *abundance*, and especially the abundance of light, entertaining, and often unreliable information that had become instantaneously available to millions via the Web. Within this new information ecosystem, many of the traditional resources for scholarship languished in paper form, out of public view; those needing them could not find them through the prevalent means of the day – that is, Web search. As such, it also makes sense that GBLP sought to serve scholarly users first, since *that* had become the public information access gap most in need of filling. To some extent, one could say that the millions of contributors of content to the World Wide Web had already built the popular piece of the collection, and Google Books was more like the second phase. As Sergey Brin told *The New Yorker* in 2007, “comprehensiveness isn’t just about, you know, total number of words or bytes, or whatnot. But it’s about having the really high-quality information. You have thousands of years of human knowledge, and probably the highest-quality knowledge is captured in books. So not having that – it’s just too big an omission” (Toobin 2007).

3.1.2.3 *The Genealogist*

Finally, there is one other favored user group that merits mention here, which does not seem to have been explicitly anticipated or hoped for by any of the project’s leadership, but which turned out to be especially well-served by its products: that is, genealogists, and amateur historians more broadly. In some ways, these users fall between the two groups already discussed: more serious about their research than the average public consumer, but

less solidly within the scholarly mainstream than the average professor or student. These users are never mentioned in any of the early public statements from either Google or its library partners, yet came up frequently in the interviews. One Googler described the project's appeal to this group especially enthusiastically:

it's really really popular amongst genealogists. And I can understand why. And if you search under... if you do any kind of a vanity search on your name, or, if your name is at all slightly unusual, like my last name is [X], which isn't that... you know, widely held. You just see all these references to ancestors and it gets really really interesting. [...] So, um... and that's something I think that full text really gives you. That when your ancestor is mentioned in like one paragraph of a book, it's not gonna show up in the card catalog, right? You're only going to find it in the full text. So we get a lot of that, we get a lot of history... um... people who are interested in history... and I guess that makes sense given what we have full view (P12).

Though those involved did not specifically foresee how useful full text access to historical works might be among genealogists and their ilk, they were thrilled to discover this was the case – at least at Google (a similarly energetic description was offered by the other Googler interviewed). Among librarians, the tone was a bit different. One seemed quite dismissive of Google's enthusiasm in this area, commenting, *"I mean my sense was they were thrilled about genealogy, they were thrilled about local history, they were thrilled... the anecdotes that they promoted were pretty lowbrow"* (P9), though others seemed more accepting. As one speculates:

I'm willing to bet that genealogists love Google Books. As a way of pursuing that sort of particular, popular, intensive research interest. Where there are in fact a number of specialized databases that they need access to. But the fact that Google, or Google Books rather, has things like county histories and stuff like that. Make it... I would imagine an extremely helpful tool for that kind of historical inquiry (P13).

For research librarians, as for archivists, there is a lengthy history of animosity toward and stereotyping of genealogists, even though – or perhaps because – they are some of the most avid users of historical collections (e.g., Redmann 1993, Kramer 2012).¹⁴⁶ And while that attitude may be abating within the library community, it's still the case that genealogists are not the most privileged library user group. Yet, among both Googlers and librarians, there is strong recognition of the GBLP's utility for these users, and an overall sense of pride in the fact that it is fulfilling their particular needs.

3.1.3 Not-so-Favored Users

At the inverse end of the spectrum from the types of users identified above, there are a few types of potential user that were identified as being especially poorly served by Google Books.

¹⁴⁶ Indeed, I have personally observed this attitude repeatedly in my time both as library staff and as a researcher working in various archives.

The main such user type – those without access to the Internet – would clearly have difficulty using *any* digital book collection; this limitation is far from limited to Google Books. Still, particularly given the project’s universalist rhetoric, it does bear mentioning that there is a large portion of the global population that the service does not and cannot reach, essentially coterminous with being on the underserved side of the “digital divide.” A few participants raised this issue quite explicitly, noting, for example, that “*obviously people who don’t have access to online, it doesn’t serve well at all*” (P10), and that use would be harder for “*the people who don’t have access to technology*” (P18). Still, P18 immediately added an important qualification, explaining,

But they’re no worse off than they are now. Because... and the physical copies will always be there. This isn’t... you know, it’s a digital copy, they are not destroying the original artifacts. But over time that... you know the digital gap is... narrows. And more and more people have access to it (P18).

The point is essentially that although digitization cannot serve those without access to digital technologies, it still allows information to flow to vastly more people than it could have reached had it stayed within the walls of the library; put differently, millions more people can afford access to the internet than can afford a plane ticket to Ann Arbor or Stanford (Jones 2006). And of course, it is worth noting that a piece of this digital divide in access to GBLP resources would have been chipped away at least a tiny bit by the settlement agreement, through the provision placing public access terminals in public libraries that would have provided full-text access to both public domain books and books that were in copyright, but not commercially available (Band 2008, 7-8). But that provision of the settlement was subject to a wide variety of often valid criticism (e.g., Albanese 2008e, American Library Association, et al. 2009, Coyle 2009, Suber 2009), and in any case, the settlement was not ultimately approved. Thus, although public library users can still access the public version of Google Books (and thus all the public domain books in the corpus) at whatever Internet-connected computers their local library may have available, their access is not nearly as broad as it might have been under the settlement.¹⁴⁷

The second segment of underserved potential users – that is, those outside the United States – arises more explicitly from the settlement itself, and the discussions surrounding it. Although an explicitly international, even global audience had been described by many participants early on in the project (as cited repeatedly above), the provisions of the settlement agreement, as well as its revision, changed the discourse in this area. Without going into too much tedious detail, two elements of the settlement are important here. First, because it was a settlement of a dispute between U.S.-based parties on a matter of U.S. law (that is, U.S. copyright), its provisions – including extended access to the book scans – only ever applied to the United States, and to the services provided within its boundaries (Band

¹⁴⁷ But then, that is true for us all.

2010, 264). And second, because many international authors and publishers had taken issue with the original settlement agreement, the amended settlement limited its scope to exclude most works of foreign origin (Band 2011, 14). Alongside the announcement of these provisions, the rhetoric surrounding the project as a whole shifted dramatically, replacing all the earlier references to globalism and international reach with a much narrower focus on greater access to knowledge *for Americans*. For example, in a joint university press release regarding the proposed settlement, UM Librarian Paul Courant is quoted as saying, “It will now be possible, even easy, for anyone to access these great collections from anywhere in the United States” (Cunningham 2008) – a significant scale-back since 2004, when the UM Libraries cited the desire to open up their collections “to users throughout the world” as a primary motivation for becoming involved in the project in the first place (University of Michigan 2004b).

Moreover, the issue of international reach remains an issue even beyond the settlement’s failure. Google is much more conservative about what it counts as being presumptively in the public domain in countries outside the U.S.; indeed, its assessments are often much more conservative than those of its international partners. As one Oxford librarian explained,

So we initially started off thinking we would be able to scan everything up to 1900 and then that date began to be brought back really at the behest of Google themselves, who began to take an increasingly cautious view. And so that removed a whole lot of material from the scope of the project. [... We] had come to a view ourselves of what the risk was of material that was... essentially, could have been published by somebody who wrote the work very young and then lived to be very old, then 70 years on top... and we thought that 1900 was a safe date and Google began to be increasingly concerned that this might not be the case. And so they pulled it back to the middle of the 1880s (P16).

And of course, this is just for *scanning* – access is another thing entirely. Though Google does not explicitly state anywhere on its website what the date cutoff for international full-view of international works might be, HathiTrust, which was largely built on Google-scanned works, does offer such a statement – and the cutoff date is 1873 (HathiTrust Digital Library n.d.-a). And although it is difficult to conclusively confirm from where I sit in Ann Arbor, I have heard from international colleagues that the cutoff is similar when using the Google Books interface. A grant-funded project at the University of Michigan, CRMS-World,¹⁴⁸ is working on making more of the works in HathiTrust published internationally between 1873 and 1941 more broadly available, but even under that project, works that are still in copyright in the U.S. all remain closed in all jurisdictions, even if they are in the public domain in their country of authorship, because to do otherwise would simply make things too complex. As the project documentation explains,

¹⁴⁸ CRMS stands for “Copyright Review Management System,” as mentioned in an earlier footnote.

Although we would like to be able to make foreign works available to foreign users, we do not have the technological or financial resources to do so at this time. Adding formal HathiTrust partners in other nations may allow us to develop a solution to this issue (Glushko 2012, 6).

As one New Zealander related to me at the 2012 ALISE conference,¹⁴⁹ such restrictions on international access can be especially frustrating when a non-U.S. user is attempting to make use of a piece of his or her own local cultural heritage, which seems clearly to be in the public domain as locally defined, but seems to be held hostage by this massive American company. And yet, though I do not wish to downplay this frustration – it would bother me hugely as well – it is worth noting that in many cases, the only reason that international user knows that that locally-published book exists is because, although it may not be fully readable in Google Books, that project has scanned it and made it searchable, and thereby opened it for discovery. There is no disputing that the service does not offer access to as many full text books in countries outside the U.S. as it does within U.S. borders; however, this is arguably almost entirely due to the massive added burden of establishing the term and expiration date of copyrights that are linked inextricably to author death dates, as they are in most of the world (including, since 1976, the U.S.) – and not to any particular failing on the part of libraries or Google themselves.

3.2 Collections

Just as the Google Books Library Project comes the closest to serving *everyone* of any of the initiatives examined here, it also comes closest to gathering together *everything*. Although upon reflection, several interviewees concluded that true comprehensiveness for a project like this – that is, scanning every book ever published, worldwide – was unfeasible, the rhetoric of the project, as well as the actual vastness of the collection it *has* scanned, indicate the significant extent to which this project was geared toward getting as close to *everything* as humanly possible.

This section will comprise three parts, examining first, the size of Google’s book collection, as it has grown over time; second, the meaning and possibility of “comprehensiveness” where this project is concerned; and finally, what sorts of things the collection contains, and where its compositional boundaries lie.

3.2.1 Size

One of the most closely controlled facts regarding the Google Books Library Project throughout its existence has been the running total of volumes it has scanned. Both in my own work on the project in its early days and later in the interviews for this project, the concept of “the public number” came up frequently, referring to the number of volumes

¹⁴⁹ Because this is what happens when you study Google Books. Folks want to tell you every problem they have with it, pretty much all the time.

that the public was allowed to know that Google had scanned. This number was, without exception, lower than the actual number; a baseline for a “more than” statement, like the “More than X billion served!” signs on old McDonald’s restaurants. And especially in the early years, figures were only released when they were pleasingly round: more than 1 million scanned! More than 5 million! More than 10!

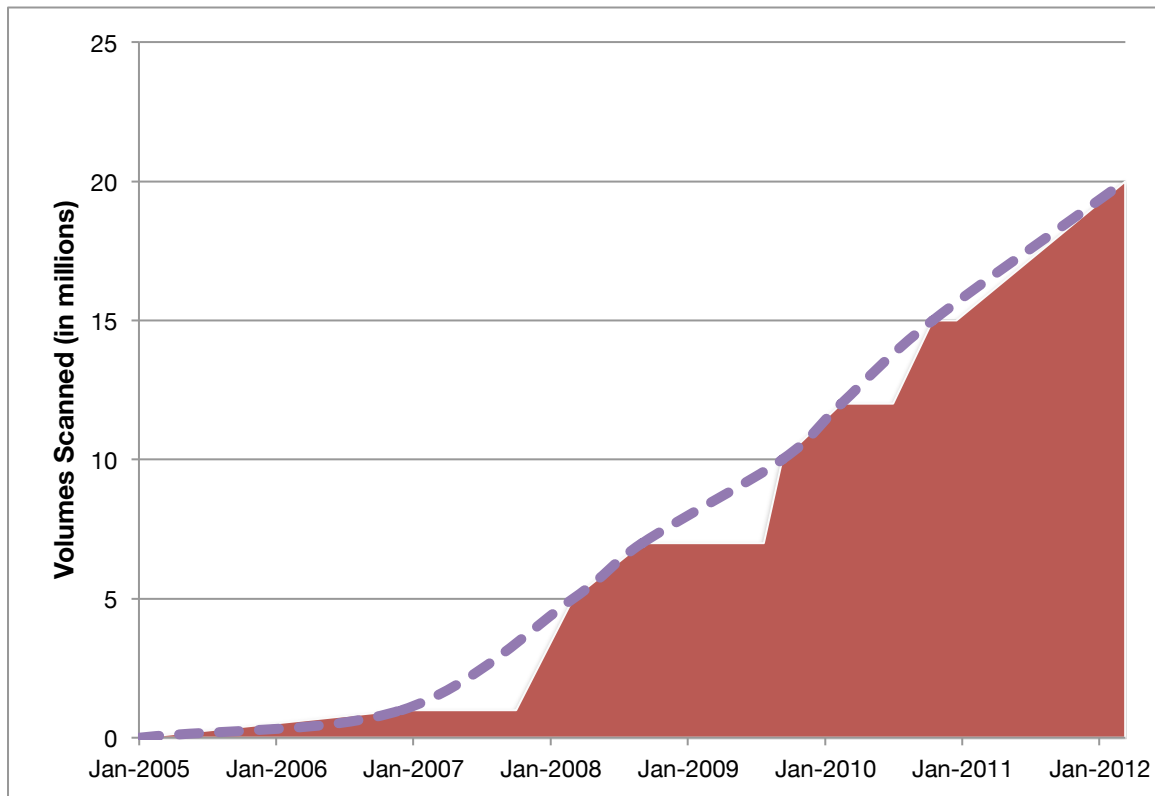


Figure 36: Public Figures for Google Books Volume Count (2005-2012)

This punctuated pattern of information release accounts for the relatively lumpy shape of the area graph (red) in Figure 36, which depicts the actual public numbers cited in Google statements and press accounts between 2005 and 2012.¹⁵⁰ As it illustrates, once a number is given – say, “1 million” – the public coverage of the project tends to stick with that number for a while – sometimes years – until Google releases the next number. It is reasonable to assume, however, that the actual trend of Google’s scanning progress has been a bit more linear; something like the dashed purple line, which uses the first mention of each successive “public number” as its only data points, and then interpolates the rest.

¹⁵⁰ Chart data: (P4, P12, Holahan 2007, Milliot 2007, Quint 2007, Read 2007, Reinstein 2007, Jacsó 2008, Leetaru 2008, Stripling 2008, Badger 2009, JTW 2009, Mahoney 2009, Otake 2009, Sands 2009, Albanese 2010c, Crawford 2010, Murray 2010, Oder 2010, Orwant 2010a, Orwant 2010b, Pon 2010, Banerji 2011, Chen 2011, Jung 2011, Howard 2012, Darnton 2013).

Another interesting issue – though one on which reliable data are even more difficult to find – is the relative size of each library’s contribution to this collection. From the data that are available, it seems likely that the largest contributor to the collection thus far has been the University of Michigan, which started off by committing all 7 million volumes in its collection and has stuck by that number consistently over time (e.g., Quint 2004, University of Michigan 2004a, Last 2007, University of Michigan n.d.-b). Two other entities – the CIC and Stanford – initially declared their openness to contributing greater numbers than Michigan (10 million and 8 million, respectively); however, it seems unlikely that either institution has done or will do so. An interviewee from the CIC explained that 10 million was more or less a fabricated round number, rather than an actual planned goal,¹⁵¹ while at Stanford, the project’s legal issues led to a contraction in the scope of the school’s contribution, to include only older, less legally risky works (P9, P17, Quint 2004, Committee on Institutional Cooperation and Google 2007, §2.1). The University of California committed the next-largest amount, at 2.5 million (University of California and Google 2006),¹⁵² and five other partners committed to contributing about one million volumes each (Carr 2005b, Barnett and Willmann 2007, Cliatt 2007, Mattiuzzo 2010, Bayerische Staatsbibliothek n.d.). The other fifteen partners for which I have data¹⁵³ all appear to have committed to contribute fewer than one million volumes – and generally fewer than 500,000 (“La Literatura Catalana” 2007, “La BCU De Lausanne Partenaire” 2007, Anderson 2007, Columbia University Libraries 2007a, Keio University 2007, Marsh 2007, Universiteitsbibliotheek Gent 2007, Van Gemert 2009, Colombet 2010, Weisl 2010, British Library 2011, Weisl 2011, “Búsqueda De Libros”).¹⁵⁴ Figure 37 provides a rough sense of the breakdown, with all the caveats noted above (as well as in the footnotes). The available data here is undeniably patchy and far from adequate, and this chart should be taken with more than a few grains of salt. Still, I have opted to include it because, if nothing else, I believe it gives a useful sense of each participating institution’s enthusiasm for the project. The bars are color coded to reflect my assessment of the reliability of the data on which they are

¹⁵¹ Additionally, the CIC is a consortium of 12 universities; as such, “up to 10 million” from the consortium represents 1 million from each individual institution (leaving aside the University of Michigan and the University of Wisconsin, who had their own agreements prior to the CIC’s involvement).

¹⁵² Although the University of Virginia declared their entire 5.1-million-volume collection fair game for scanning (University of Virginia Library 2006), I was not able to locate any further data suggesting how many out of that 5.1 million volumes they actually intended to scan (or actually did scan).

¹⁵³ The only institution for which I could find no number cited anywhere was the New York Public Library; however, their catalog lists 32,000 volumes scanned through Google (<http://catalog.nypl.org/search/XGoogle+Books+Library+Project>); even if this is not a comprehensive list, it seems safe to assert that their contribution was in the thousands, rather than anywhere close to a million.

¹⁵⁴ Three institutions – Columbia, Ghent, and Madrid – only go so far as to state that they will be contributing “hundreds of thousands” of volumes; although there is a small chance this could extend beyond a million, that chance does, in fact, seem quite small.

based.¹⁵⁵

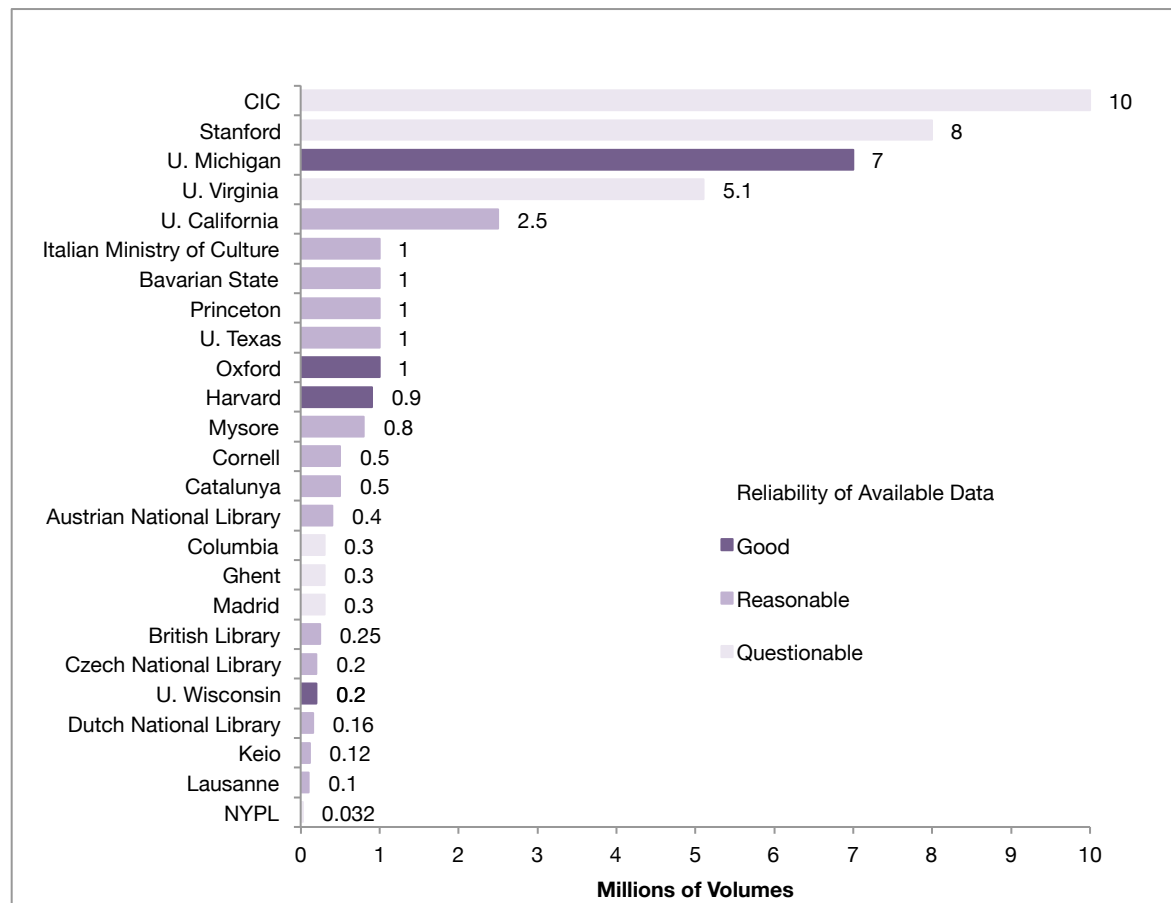


Figure 37: Intended Size of Contribution by Google Books Partner Libraries

3.2.2 “Comprehensiveness”

Beyond the actual figures, however, lies a more basic question. Namely, what would it take for a project like this one to amass a “comprehensive” collection of all the books in the world? Is such a thing possible? I posed these questions directly to every person I interviewed, and virtually every one¹⁵⁶ expressed the opinion that true comprehensiveness

¹⁵⁵ Category definitions:

- **Good** = I have some confirmation that this was or will be close to the number scanned.
- **Reasonable** = The figure was reported by a valid source (e.g. a press release or blog post from Google or the library itself), and I have no specific reason to believe it to be inaccurate.
- **Questionable** = I have reason to believe that this is *not* an accurate figure (this includes four cases where no specific number was reported: NYPL and the three institutions that only said “hundreds of thousands;” for the latter group, 300,000 was used as a proxy figure for the purposes of this chart).

¹⁵⁶ The exception being an informant from the NYPL, who answered: “Yeah, because reality is it’s a finite number. So at some point, certainly they’re digitizing a fixed amount of works. Anything that’s been published after a certain date is in digital form somewhere. So they’re looking at something... picking an arbitrary date of the year 2000 and prior. And a lot of the works are... you know it’s heavily weighted towards the end of all of that... so, um, I think it’s achievable. I think it’s achievable to an

was more of a rhetorical aspiration than a realistic objective for book digitization, generally stating flat out that it would be impossible for this project to get to every book (P2, P3, P4, P8, P9, P10, P12, P13, P16, P17). One librarian, after stating that comprehensiveness was “almost certainly not” possible, immediately added that he didn’t “think that was ever the goal of Google Books,” explaining:

The great thing about Google is that they’re engineers, and so if the last 10% are really expensive compared to the first 90, they do the 90. So... so it would be... to make accessible a WHOLE LOT of the books of the books in the world, that would be much closer I think to their original goal. I actually think there, we might, in time, through various initiatives of which this is one, we might in time get there. It’s, it’s... and this is where it’d be crazy otherwise... it’s so easy to do. And, what Google did was showed us it was easy to do rather than hard. Easy to do 90. Not easy to do 100 (P10).

This essential point – that it can be relatively easy to scan *most* of the world’s books, but near-impossible to scan *all* of them – is strongly echoed by one of the Googlers I spoke with, and crops up in some of the other librarians’ statements as well (P4, P8, P9).

In terms of reasoning for this perception of impossibility, two themes emerged within the interviews – one legal, one ontological. The legal reasoning is fairly straightforward, and will be discussed in a bit more detail in the next section; here it will be enough to state that opt-outs by copyright holders (especially major journal publishers (P10)) effectively prevented any library willing to scan its in-copyright collections with Google from doing so in a comprehensive way – and most libraries were averse to scanning beyond the public domain in the first place (P3, P10, P13).

The ontological issues are more abstract, and to my mind much more intriguing. That is, when asked about whether this project could ever be comprehensive, many interviewees immediately raised much broader questions: one immediately shot back “Define ‘book’” (P3), and another elaborated that,

the short answer [regarding comprehensive scanning] is no. The slightly longer answer involves asking what do we mean by book? And what really do we mean when we say, “everything”? You know. Everything that has been published in formally published in codex form, somehow? Or are we talking about everything as manuscripts, or maps, or things that are not per se books, but involve words somehow (P13).

Building on the latter question, one Googler mused that library collection and preservation decisions create a further roadblock to comprehensiveness (one cannot get all the books ever published, because not all the books have been saved), and that adding more and more smaller and smaller pieces would ultimately garner diminishing returns:

the collective library system, of every library in the world, is not comprehensive. Each

organization like Google that thinks big and has deep pockets. They continue to sign up new institutions. They continue to enlarge the type of material they have. So over time Google will have everything, but right now I’m sure there’s gaps (P18).

library makes decisions, that is not comprehensive. As we say...ultimately, the libraries have made decisions about what to preserve, um, it is not necessarily the case that all those decisions were the right decisions. Right? Um, now the good news is, there's overlap. Right? And so collectively, and so um, uh, when you say what is comprehensive, there is no...you're never comprehensive. [...] So, um, I don't think anyone can build even a list of all the books that've ever been published. And a list of all the books that're held in some...thing that looks like a library. Right? So I do think in the end you're talking about resources, and you're talking about impact. And it is the fact that um, uh, in terms of the impact, um, uh, y'know, there's a huge, steep curve, even with libraries, and that curve keeps going and going and going and going and going (P4).

This, in fact, is a truly wicked problem: how would one even go about identifying and locating “all the books in the world,” much less scanning them all? What counts as a book, and what does not?¹⁵⁷ One Googler took a run at these questions in a 2010 blog post, algorithmically crunching numbers from more than 150 metadata providers to arrive at a figure just shy of 130 million books (Taycher 2010) – but given the questionable quality of the data Google receives from those metadata providers (e.g., Nunberg 2009a), this figure seems highly suspect as well – and even were it not, *counting* all the books is very, very far from *finding* all the books. After all, “all the books in the world” are likely spread quite unevenly all across the globe; pretty soon one would get to the point where the next volume to scan only exists in one small village in central Mongolia (or worse, if you're a U.S. company, Iran or North Korea), and the cost/benefit ratio for scanning that volume dives straight through the floor.¹⁵⁸

Still, shifting back from impossibilities to opportunities, several informants also suggested that although truly universal comprehensiveness may be a pipe dream, this project (and others like it) should certainly be able to achieve a critical mass beyond which most of its users would be able to find any book they could possibly want – essentially, a state nearly indistinguishable from comprehensive coverage at the user level (P4, P8, P9, P12). As one librarian explains,

if you have 20 million, and then, you sort of... crawl your way to 30 million... you know this is... I mean those are pretty remarkable... pretty remarkable numbers. [...] So, every book? No[...] it's possible that some things are going to get lost. That shouldn't be lost, because people assume that everything's there, when it's not. But... um, you know on the other hand I'm thinking that you know if I wanted to find something that I read in

¹⁵⁷ There are other questions nested within this, as well, which pretty quickly arrive at FRBR issues – OK, so you want every book in the world. But do you want every manifestation? Every edition? Every translation? How many versions of *Romeo and Juliet* is enough?

¹⁵⁸ This, of course, is where Google's library partnership strategy comes in: start with some really large and wide-ranging collections (e.g. Michigan, Stanford, California), then try to fill in the remaining gaps those collections leave by seeking out distinctive collections in other places. Eventually (and not too far in, really) there comes a point where it's hard to see (a) where the remaining gaps are, and (b) who the next partner would have to be in order to fill those gaps. (And beyond that, not every library in the world is even willing to partner with Google...so there's that.)

graduate school I'm guessin' Google's touched it already. Just... just sayin' (P9).

"True universality," another librarian noted, "*is chimeric*" (P2). Still, those involved in Google Books professed a strong faith in the project's capacity to digitally amass not *everything*, but a very significant segment of the world's books; as one Googler put it, to make "*a serious dent in the pile*" (P12).

3.2.3 Composition

We have established that the digital book collection created by the Google Books Library Project is extremely large – more than 20 million volumes, and still growing – but will almost certainly never encompass every book ever written. So what do they have? Or perhaps more to the point, what do they *not* have, and what are they *unlikely to ever get*? This section will address each of these questions in turn, focusing primarily on the latter.

3.2.3.1 Coverage

In order to get a rough idea of what the GBLP corpus contains, one must again recall where the raw materials to build that corpus have come from – that is, a set of major research and governmental libraries.¹⁵⁹ Such libraries tend to be oriented more toward scholarly works than toward, say, popular fiction. As one Googler suggested when asked about the strengths and weaknesses of Google's collection,

you can look at our current partners and that describes, and you can see from there, a certain distribution of types of institutions. There are public libraries in there, but the type of stuff that public libraries hold often are different, so...um, I, academic libraries tend not to be as good at preserving, y'know, certain fiction, romance, fantasy, all sorts of things. One statement that Paul Courant once said to me, he says, "I doubt Michigan has a copy of all of John Grisham's novels" (P4).

As he suggests, research libraries tend not to collect much popular work, and when they do, they don't always hold onto it over time.¹⁶⁰ And of course because many libraries opted only to allow Google to scan their public domain materials, the GBLP corpus will tend to be weighted toward older works – items published before 1923 in the U.S., or before about 1870 internationally.

Still, also largely thanks to the collecting priorities of its library partners, the GBLP corpus contains works in hundreds of languages: a Google blog posting puts the number above 400

¹⁵⁹ The Google Book Search corpus, as distinct from the Google Books Library Project corpus, also contains works from tens of thousands of publisher-partners; the materials Google has collected through those partnerships are undoubtedly quite different in scope, content, and age from those it has gathered from its library partners. As in the rest of this chapter, the focus here will remain on the library end of the project.

¹⁶⁰ Indeed, even though public libraries do collect things like paperback mysteries and romances, they *also* tend to weed those parts of their collections fairly ruthlessly over time. While a public library may well have a full set of John Grisham, they may equally well *not* have every Harlequin romance, even if they had at some point purchased each one.

(Crawford 2010), and that figure is borne out through examination of the languages available in HathiTrust – 427 – whose content comes predominantly from Google Books (HathiTrust Digital Library). Indeed, although HathiTrust does also contain books from other scanning projects, including the OCA, it relies heavily enough on Google to make a brief word on its language breakdown relevant here. In particular, although most of Google’s partner libraries are located in English-speaking countries, slightly less than half of the HathiTrust corpus is in English, with strong showings by many European languages, as well as Chinese, Japanese, and Arabic. Figure 38 illustrates this breakdown.

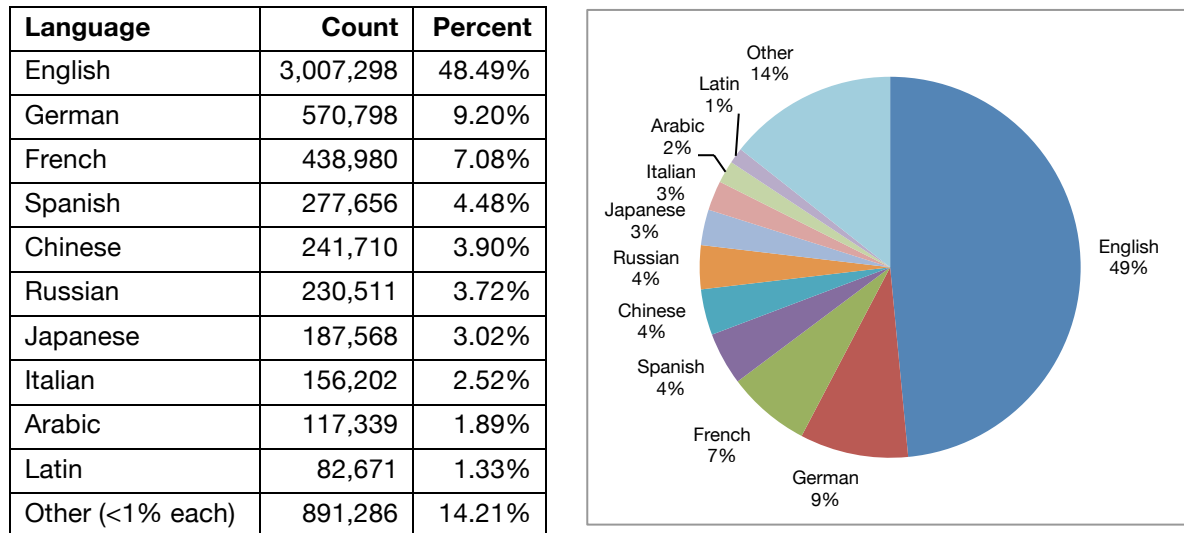


Figure 38: Language Breakdown of Volumes in HathiTrust, as of July 12, 2013

Google has also developed OCR capabilities for many languages that it previously could not digitally read or index. As one Googler explained,

So...I mean, we OCR...I know it d- y'know it had been over a hundred languages. We have, ya know, Thai, Fraktur, um, Chinese, Japanese, Korean, [...] it's extensive. With many different scripts. We now do...um...Arabic, I'm pretty sure we do Hindi, um and we spent a lot of time on that, being able to OCR (P4).

And further, he went on to assert that even if Google had not developed an OCR library for a particular language, they would sometimes scan those things anyway – and in those cases, “if you speak [the language], you can’t search it, but you can read it” (P4). Still, at least early on, some libraries saw lack of OCR capabilities as a barrier to what they could scan. Specifically, an informant from Harvard asserts,

I would say that part of why they [Google] went into developing lots of OCR techniques for random languages is because we made so much noise about it. Because that was part of the shelf-pulling problem. Where if they did not have OCR’ing capacity for late nineteenth century Arabic, why send it? But then if we don’t send it, we have to manage how things around it were handled. So it was really interesting (P13).

It seems like in this area there was another chicken-egg problem: Google couldn't OCR it, so Harvard didn't want to send it – but if Harvard had sent it, perhaps Google could have figured out a way to OCR it. None of the other institutions represented in my interviews raised this as an issue, though it is perfectly possible that such issues occurred elsewhere as well, particularly at institutions that were more conservative about what they were willing to send.

Overall, the GBLP corpus would seem likely to represent a very good general scholarly collection, in many languages, at least for works published prior to 1923 (or by the U.S. Government, or falling into some other, similar, gap in the copyright wall), and a reasonably good tool for finding (though not *reading*) work published after that (via snippets and “No Preview” metadata hits in their search results, as well as things included through the Partner Program). It would most likely be less useful as a source of free leisure reading, since Google's *free* books tend to be fairly academic, and Google's *leisure* books tend to be for pay (i.e. Partner Program books) – and in that, its collection differs significantly from those of physical public libraries, at least in the present day.¹⁶¹ Additionally, many items have been excluded from this collection for reasons entirely unrelated to content, which has made the gaps in it quite unpredictable and idiosyncratic. These limitations and exclusions will form the final piece of this section.

3.2.3.2 *Limitations*

As with many aspects of the Google Books Library Project, there was a fair amount of variation in the limitations placed on scanning among the different partner libraries. And although much of this variation emanated from differences in library risk-aversion and willingness to experiment, some also has roots in the differential balances of power between Google and each of its library partners. Specifically, while some libraries (e.g. Michigan, Stanford, Harvard) appear to have had quite a bit of influence in deciding which books would be included or set aside, others – mainly later partners – appear to have had less (P3, P8, P10, P17, University of Michigan and Google 2004, Committee on Institutional Cooperation and Google 2007, University of Texas and Google 2007). This imbalance is evident in the available scanning contracts: for example, Michigan's contract states loosely that “*The parties shall cooperate to identify Available Content to be Digitized,*” while the CIC's, signed just a few years later, declares that “*Google reserves final discretion over which Available Content it will Digitize*” – quite a shift in power dynamics within a short span of years (University of Michigan and Google 2004, Committee on Institutional Cooperation and Google 2007, emphasis added). Still, although their manifestations may have differed, there were three consistent factors that led to the exclusion of materials from Google's

¹⁶¹ Of course, if one wished to read their way through the *classics* for leisure – things like Shakespeare, Cervantes, Dickens, Dumas, and so on – Google Books is an absolute treasure trove.

scanning project across all its library partners: (1) legal issues, (2) format constraints, and (3) avoiding duplication.

3.2.3.2.1 *Legal Issues*

Out of Google's 25 library partnership agreements, only eight included provisions for scanning works still covered under copyright, and all of them were signed between 2004 and 2007, before the Settlement Agreement was first announced. The other 17 agreements, including all of the agreements with institutions outside the United States, only extended to scanning items in the public domain. Figure 39 provides a breakdown of these agreements by type and year.¹⁶²

Public Domain Only	Both PD and In-Copyright
NYPL	U. Michigan
Harvard	Stanford
Oxford	U. California
Madrid	U. Wisconsin
Catalunya	U. Virginia
Princeton	U. Texas
Bavarian State	CIC
Lausanne	Cornell
Mysore	
Ghent	
Keio	
Columbia	
Italian Ministry of Culture	
Austrian National Library	
Dutch National Library	
Czech National Library	
British Library	

Partnership Year
2004
2006
2007
2010
2011

Figure 39: Breakdown of Copyright Decisions by Google Library Partner and Year

Even among those institutions who did agree to scan in-copyright materials, moreover, it is unclear to what extent this was actually done – though most indications would suggest that it was less than intended. For example, the University of Wisconsin signed on with a plan to start with public domain materials and then move into in-copyright collections at a later date (University of Wisconsin 2006); it is not clear whether that shift ever occurred. And Stanford, which had initially expressed a strong commitment to scanning its entire collection, ended up feeling pressed to scale back its ambitions due to legal concerns – and particularly the fact that, as a wealthy, private institution, it could make an appealing target for copyright litigation parallel to that pursued against Google. Some in-copyright items were still scanned at Stanford, but far from all of them, and Stanford does not have access to copies of those that have been scanned; they are being held in escrow by Google until Stanford's legal counsel feels the risk of litigation has sufficiently subsided (P17). And even at institutions where in-copyright scanning definitely did occur, such as the University of Michigan and the University of California, that scanning had ceased by the time of this

¹⁶² Table data sources: (Quint 2004, University of Michigan 2004a, Carr 2005b, University of California and Google 2006, University of Wisconsin 2006, University of Wisconsin and Google 2006, "La BCU De Lausanne Partenaire" 2007, "La Literatura Catalana" 2007, Anderson 2007, Barnett and Willmann 2007, Cliatt 2007, Columbia University Libraries 2007a, Keio University 2007, Last 2007, Marsh 2007, Universiteitsbibliotheek Gent 2007, Committee on Institutional Cooperation 2008, Colombet 2010, Mattiuzzo 2010, Weisl 2010, British Library 2011, Weisl 2011, "Búsqueda De Libros", Bayerische Staatsbibliothek n.d., University of Michigan n.d.-b)

☐ Scannable
☒ Unscannable

Details

- ☐ Poor condition
- ☐ Oversized Book
- ☐ Uncut pages
- ☐ Publication date
- ☐ No Metadata
- ☐ Unsupported Format
- ☒ Opt Out
- ☐ Duplicate Book
- ☐ Partner Request
- ☐ Previously Scanned – DO NOT MARK UNSCANNABLE IN ISC

Book Condition

IJC	IJD	
TBD		
BP2	BP3	BP4
TP2	TP3	TP4
BBC		

The scanning operator will notify a lead that an **UNSCANNABLE** condition has been found.

The lead will insert this white flag inside the book and initial at the top.

Figure 40: Slip found in a book at the University of Michigan's Hatcher Graduate Library, indicating that the book had been opted out of scanning by a rights-holder.

writing, leaving many in-copyright items un-scanned, and unlikely to ever be scanned under the aegis of this project (P10, follow-up conversation, July 12, 2013).

Risk aversion, moreover, is not the only legal issue that ended up constraining the Google Books Library Project's scanning: there was also the issue of opt-outs. One of the most controversial pieces of the project from the beginning, the opt-out system was the way in which Google handled cases where copyright holders did not want their books included in the index. Analogous to the "no robots" files that can be attached to webpages to keep search engines from crawling them, the opt out system provided a means for authors and publishers to tell Google not to crawl their books (Helm 2005a).

Initially, Google had planned to scan everything the libraries would let them, regardless of opt-out status, and then exclude particular scans from the publicly searchable index if their owners opted them out. After the authors and publishers began threatening litigation, however, Google announced that it would shift its opt-out policy to extend to the scanning itself – that is, an opt-out would not just prevent a book from being indexed, but also from being scanned in the first place (Albanese 2005b). For the libraries hoping to have their in-copyright collections scanned, this was a major blow, as many publishers opted out their entire backfiles, wholesale (P10). Clearly, the publishers were well within their rights in doing this; however, it does represent a significant limitation on the GBLP collection, as well as a blow to the comprehensiveness of scanning at individual institutions – especially Michigan and Stanford, where the goal had been to procure backup copies of entire collections (P10, P17). And although many of these same publishers have contributed partially-viewable versions

of many of their books to GBS via the partner program, it is unclear whether they have put forward as much of their own backfiles as the libraries would have been able to – and of course such contributions do not help the libraries fulfill their objectives for the project, since they only get copies of items actually scanned from their collections. Figure 40 – a slip

found by happenstance while browsing the University of Michigan Library stacks – illustrates how one library marked books as being opted out (and, in the surrounding list, also shows several other criteria on whose basis items could have been, and likely were, excluded from the Google scanning).

3.2.3.2.2 *Format Constraints*

Even among items that were not subject to opt-out or hesitations based on legal risk, other limitations still held force – and the most powerful of these were the pragmatic, physical constraints of Google’s scanning equipment. Research libraries, after all, hold many items that are not traditionally book-shaped¹⁶³ – or to put it more technically, in *codex form* – and even among codices, there is wide variation in size, legibility, and condition. In order for Google to scan non-destructively using the scanning machinery it had developed, constraints were implemented in terms of:

- Size (not too big, not too small)
- Medium (traditional codex; not unbound materials, microfilm, newspapers, maps etc.)
- Condition (able to withstand the scanning process without falling apart)

These constraints are cited throughout the interviews, as well as in several public descriptions of the project (P4, P10, P12, P13, P16, P17, P18, University of Michigan 2005, Stanford University Libraries 2006, Ghent University Library 2009, Princeton University Library 2010). The first two criteria – size and format – seem to have been broadly consistent in implementation across partner institutions. There was more variation, however, in judgments regarding condition, with Oxford in particular cited as being more willing to allow Google to scan older and rarer materials than its U.S. library counterparts (P12, P16).

Interestingly, because this set of exclusion criteria have purely to do with the physical form of the books, and nothing to do with their content, it is difficult to say what cumulative effect such exclusions might have had on the shape of the collection. (Are books on certain topics likely to be more fragile? Larger? Smaller? More valuable?) Extrapolating a bit from the criteria, one might imagine that art and architecture books would be at a disadvantage, since large folios are especially common in those areas (they also include a lot of images that would be difficult to clear rights for, but that’s a separate issue). And it’s possible that although the public domain focus of the collection biases it toward older works, the condition standards tilt it away from the very oldest. Overall, however, it is extremely difficult to tell what sorts of content – what actual swaths of information – will be missing

¹⁶³ For example, I remember loving to wander through the aisles of scrolls in the University of Chicago’s Regenstein Library stacks as an undergraduate, admiring their bright yellow tassels and wondering how they managed to stay intact in such openly accessible storage. Despite all their beauty, intrigue, and likely historical importance, however, such collections are prime examples of the sort of materials universally excluded from Google’s scanning.

from Google's book corpus because the container it came in was the wrong size, shape, or condition. When asked about where the gaps would be, when all was said and done, one librarian bluntly replied, "*I think it's going to be random*" (P8) – and given the issues just described, I would tend to concur with his assessment.

3.2.3.2.3 Duplication

The final limitation on Google's scanning – the company's growing concern with avoiding duplication as the project progressed – is less a limitation on the comprehensiveness of Google's own collection than an additional hole-punch in the aspirations of its library partners to gain complete backups of their own collections. Early on, Google's representatives (and especially Larry Page) claimed left and right that they didn't care about duplication. They would just do shelf-clearing – scan all the books and sort out the duplicates later (P9, P10, P17). However, as the project got underway, they soon changed their tune. A few years in, the company began to turn away books that their system tagged as already having been scanned from another library (P10, P17). As one Stanford librarian describes,

once they got the metadata figured out in a much more precise way they could determine that they had already scanned something. And they, because they felt they had a good capture, a good scan, rather than in the past where they weren't so sure, they didn't have to do it a second time. So they reject things they already felt that they had. So again, that's a milestone for them, but that's what made it then they wouldn't be scanning our copy and we wouldn't get our copy (P17).

In the interviews, librarians at both Stanford and Michigan noted problems with this shift in procedure: first, the metadata was not, in fact, quite good enough, and would frequently erroneously turn away alternate editions of, say, *Robinson Crusoe* as duplicates (P8), and second, it further eroded Google's initial commitment to scan each of those libraries' entire collections (P10, P17). Although it represented a reasonably clear gain in efficiency for Google, given Google's motives and views on the book sphere (it's all a bunch of data, and what's the point of duplicate data), it was an equally clear loss for the two largest and earliest library partners, who would now have essentially completely random holes in their digital collections where they hadn't managed to get a particular book to Google before anyone else did.¹⁶⁴ As one Michigan librarian laments,

We're way more than half done. But we're a good ways from being done. And what's done mean when the size of the collection to be done has changed? The other thing's, originally Google wasn't going to worry about duplicates and now they're worried about duplicates. [...] So we're, we're never going to be... we're never going to get our collection done (P10).

¹⁶⁴ In a sense, it made it into a kind of race, where whichever library sent a book first got the prize of a digital copy, and all the rest did not.

Ultimately, the issue of duplication became a significant point at which the interests of Google and its two earliest and largest library partners, in other ways well aligned, came to widely diverge. However, it bears emphasizing that this divergence is only truly significant at these two institutions; all of the other library partners were made aware from the outset that Google would scan only a portion of their collections, and in all of the later public contracts and statements, it is quite clear that Google had become focused exclusively on scanning unique materials, and reserved the right to reject any individual item sent to it if it was considered duplicative (e.g., Marsh 2007, University of Texas and Google 2007, Committee on Institutional Cooperation 2008).

4. Implementation

As in the previous two chapters, the final section here will describe the processes and structures through which information was made available in this initiative. In this case, the level of detail will be somewhat greater, as the data gathered for this project provide an especially useful window into a set of as yet thinly understood processes.

4.1 Processes

The processes involved in order to get a digital book out to users in the Google Books Library Project are considerably different from those required to perform the same task for physical books in public libraries, though they would have some crossover on a Venn diagram. This section will focus on the processes required to move a book from stack to screen, as diagrammed in Figure 41, below. The numbered segments along the left side of the diagram correspond to the first four subsections here: (1) selection and order of submission, (2) library pre-processing, (3) transit and scanning, and (4) digital delivery. The flow chart follows the progress of an individual volume as it moves through the system.

Many details have been omitted here, and given the variation in involvement among libraries, it is likely that this flow does not fully capture the experiences of every one. It is also worth noting that Google's procedures seem to have iterated a fair bit over time, as they tried out different approaches (e.g., P8, P12); as such, this depiction is more of a snapshot in time than an always-and-forever view of how the digitization proceeded. Despite these limitations, however, I would suggest that the figure provides a useful aid to explicating a complex set of procedures.¹⁶⁵

¹⁶⁵ A much more detailed diagram of the particular workflow in the University of Michigan's Google partnership, created by UM librarian Suzanne Chapman, is available at the UM Library website (University of Michigan Libraries 2008). Beyond that chart, the citations for Figure 41 are: (P8, P9, P10, P12, P13, P16, P17, P18, University of Michigan and Google 2004, "Stanford and GBS: Statement of Support" 2005, Carr 2005b, Carr 2005c, University of Michigan 2005, Stanford University Libraries 2006, University of Wisconsin and Google 2006, Giersberg 2007, Universiteitsbibliotheek Gent 2007, University of Texas and Google 2007, Committee on Institutional Cooperation 2008, Levy 2011, Biblioteca Complutense n.d.-b, Cornell University Library n.d., University of Virginia Library n.d.-b, University of Wisconsin Libraries n.d.-a).

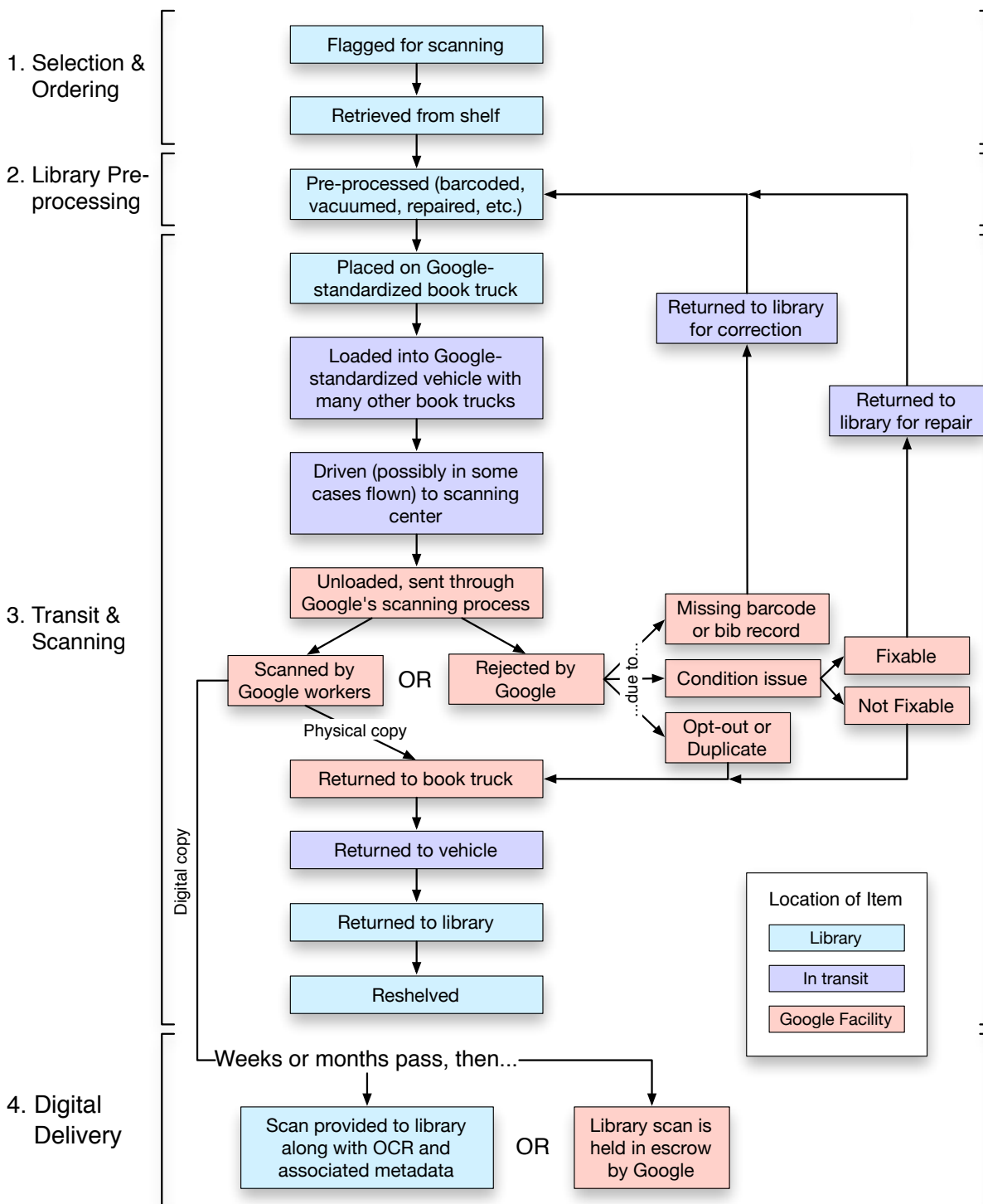


Figure 41: The Travels of a Google Library Book, from Stack to Scan (and back)

The final subsection here will deal with an element of what happens beyond the end of the diagram above: that is, the various constraints on use of the digital copies faced by (a) Google's library partners and (b) end users. Just as in the public library cases, use of

materials digitized by Google has always been subject to certain conditions and constraints: it is not fully “free,” even when the materials are offered free of charge – and this extends to both individuals and institutional partners.

4.1.1 Selection and Ordering

Like most pieces of the Google Books Library Project, the process of deciding which books to scan varied among library partners and also shifted over time. For the first five libraries, Google’s interest was in procuring as many books to scan as it possibly could; starting from zero, duplication was not yet a concern (P9, P10, P17).

As such, the decision of what and how much to scan from each of the G5 libraries rested mainly with the libraries, and came down to each one’s level of risk-aversion. The University of Michigan, its confidence bolstered by its sovereign immunity¹⁶⁶ and concurrence with Google’s fair use analysis, opted to allow Google to scan all of its books, in order of the convenience with which they could be cleared off the shelves (P8, P9, P10, University of Michigan 2005). The Stanford University Library wished to pursue an equally aggressive course, and signed on to have its entire collection scanned as well, but backed off a bit after the AAP and AG lawsuits were filed, altering the order of scanning to diminish its legal risk (public domain first, then orphan/disputed works, then in-copyright) (P17, "Stanford and GBS: Statement of Support" 2005, Stanford University Libraries 2006). Harvard, Oxford, and the New York Public Library all opted to scan in the public domain only, and also exercised varying levels of restrictiveness with regard to the age and condition of books they would allow Google to process (with Harvard being the most conservative in that regard) (P13, P16, P18, Carr 2005b).

As previously noted, beyond these first five – and even within those five in later years – however, Google began to care more about duplication. Book scanning, after all, is expensive; duplication of effort seemed increasingly wasteful. Procedurally, this manifested in two ways: first, strategic selection of additional partner libraries, and second, the development of workflows for identifying and culling duplicative materials within the scanning process, as discussed above. Beyond the first five libraries, partners were approached in an increasingly targeted way, to fill perceived gaps in the Google book collection. Negotiations with these partners thus appear to have taken a somewhat different

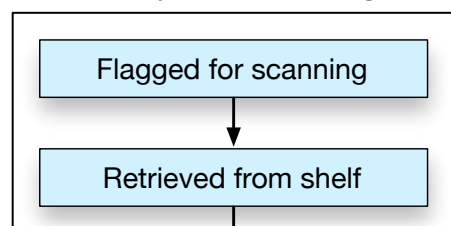


Figure 42

¹⁶⁶ “Sovereign immunity” is a legal term, referring to the concept that state institutions, as organs of sovereign governments, are generally immune from tort liability – that is, they typically cannot be sued for damages (Legal Information Institute 2010). Within the context of Google Books, this means that the liability risks are much lower for public institutions like Michigan than for private ones like Stanford; where Michigan could only have been sued to cease and desist, Stanford could have been sued both to cease and desist and to pay potentially millions in damages. (This is a rough layman’s description; I am not a lawyer.)

form: instead of Google taking as much as it could get from each library, it might specify particular collections of interest to it – and, on the other side, the libraries would generally indicate particular collections that they would like to be able to scan from end to end (for example, the CIC “collections of distinction”) (P9, P12, Cliatt 2007, Marsh 2007). These tended to be fairly complementary interests, since a library’s most distinctive collections were also likely to be more unique, and less duplicative of collections already scanned.¹⁶⁷ Still, even in these later libraries, ease of access played a significant role, alongside considerations regarding content. As one Google Library Partnerships Manager described,

the priorities have been different from one library to the next. So with some libraries we’ve had agreements to scan you know, the agreement called to scan their entire collection. Sort of everything. Start at the front door and work your way to the back door. Other libraries, we’ve had agreements and they’ve said “we don’t want to scan our entire collection. We just want to scan this particular special collection.” [...] I know that in Lyon, for example, we’re scanning the Jesuit collection. There was a Jesuit library in Paris that was closed down, and they gave their collection to the municipal library of Lyon, and Lyon was very very interested in having that particular collection scanned. [...] So, depending on sort of what the libraries are interested in, that, that affects what gets scanned in terms of prioritization...and historically a lot of the prioritization though has been about logistics than it has been about content. So there are some collections that are just sort of physically easier to get to. For the libraries. And when you’re scanning enormous numbers...high volumes, um, that comes into play a lot (P12).

The order and method by which the books were pulled also differed between library partners. Ultimately, only the University of Michigan pursued a true shelf-clearing strategy, in which books were just moved shelf by shelf to book trucks with little examination of content. As one of their librarians described, UM started with their remote shelving facility, Buhr, simply because it contained materials less likely to be missed – but as the project progressed, this ordering turned out to have unforeseen benefits for the fulfillment of the library’s objectives:

We started with storage facilities. So we started with more little-used materials, and I think if we’d started with the high-use materials and gone to the little-used materials, we’d have less of a chance of getting to everything. But we started at the wrong end, but there are good reasons to do that (P8).

As Google’s interest in the project diminished in later years, it grew less interested in truly obscure (yet often still in copyright) materials like those housed at Buhr; by starting there, UM had the opportunity to get that piece done while interest was still high.

At Stanford, the initial library cost model – the budget required for library staff to retrieve and prepare the books for Google scanning – was calculated based on the assumption that

¹⁶⁷ There are, however, some notable exceptions to this – for example, Northwestern University’s Transportation Library, which, though very important as a collection, contains a high percentage of materials duplicated at other Google partner institutions (P9).

they would pursue a straightforward shelf-clearing strategy like Michigan's. However, as legal concerns surrounding the project mounted, the university deemed that strategy too risky. Because of this, the actual order in which books were pulled at Stanford ultimately required a great deal more upfront analysis and processing – at much higher cost to the university – to scan roughly in order of legal risk. One of the leaders there describes more specifically how this process unfolded:

When we first started...I did a cost analysis of what I expected it would cost Stanford to in fact, to do this.... uh send Google all of our books. And we had approached the initial cost analysis as we were sending them everything. Which had been our agreement. [...] the notion was it would be everything, and you just take a whole chunk of a shelf and you scan the barcode and send it off. Ultimately, [...] once they got sued and our university started being nervous about what we sent them, we were limited to categories of materials that we could send them by our own University, by the General Counsel's office. And it started out with only things that were out of copyright....And, which is pre-23 in the U.S. And so we had to build our own lists of items to pull...whose publication date in the metadata was up to 1922, up through 1922. So we went through the collections and pulled everything that was pre-1922 and sent those first. And as we were approaching the end of what was going to be available, in general collections for that, I approached our University Librarian and said you know we're going to be running out in about a month if I don't get another category of material that the Counsel's office says is alright. So he went to the General Counsel, who went to the Board of Trustees, and we then got the next set of...dates that we could do. And I think it was up until or through '64. And so we went through the collection and pulled out all that stuff... Um... months and months later we had to do a... go back again... so the expense of having to go through the collection multiple times and picking book by book was dramatically higher than my original costs of I forget how much per book, I don't know, \$2 or \$3 a book. It was a lot more expensive when you had to pull one at a time. 'This book and not that book.' One at a time. (P17).

At Harvard, Oxford, and NYPL, of course, the plan had been to scan only in the public domain from the beginning, and so there was never any question of doing full shelf-clearing. Instead, books needed to be individually pulled for scanning based on their copyright status as well as other criteria (condition, format, etc.). Harvard appears to have imposed the most restrictive and labor-intensive process of any library partner involved, which ultimately resulted in a very small percentage of its collection (less than 7%) being scanned by Google:

it was felt at the highest levels that we could not risk exposure to the kinds of lawsuits that might come our way if we did the full shelf-clearing digitize everything A to Z, old Widener 1 to whatever... So, we could not, from that legal mandate from the Office of General Counsel, do this... just go in and clear shelves, and put shelves back full. It was much more complicated. We had to devise pull lists. We had to develop call it "narratives" to explain to patrons why there were these gaping holes in parts of the stacks, but not in others. It took... since we were not pulling everything from the shelves and able to work through them one by one, it involved an awful lot of post-return shelf

reading to make sure that things actually got back where they really belonged. So the technique part of the Google project in Widener and in all of the other libraries was really very involved. Yeah, took a lot, took a LOT of time. So [...] of the fourteen million volume collection at Harvard, I would say on the basis of the age of the collection, our standards for conservation issues, the size issues, you know, “too big”, “too small” to be digitized... weird languages that don’t scan well, or OCR at all... excluding serials in large part... less than a million volumes for the Harvard Collection went on to the Google project (P13).

Beyond Harvard, the process seems to have been somewhat less onerous; although there was still a perceived need to “*feed the beast*”¹⁶⁸ at places like Oxford and NYPL (P16, P11), interview descriptions of scanning order from those institutions were much looser-seeming – based more on ease of access and readiness for transport than any more precise distinctions among materials. At Oxford, for example,

It was just the ease with which we could get the material. So it was very pragmatic. A large part of our collections were shelved by size, and so we moved through them in that way. Um... before moving out into areas where it was more difficult to fetch material. So it was more about providing an even delivery of books to the scanning center rather than any more you know intellectual understanding of how to do it (P16).

Similar descriptions are given in interviews with representatives of the CIC and University of California libraries: libraries would send whatever was easiest and readiest to send at the moment at which they had to send it. At some institutions, this required copyright determinations; at others not (P3, P9). Additionally, as noted in the discussion of collections above, there was a great deal of variability in power dynamics with regard to book selection among Google’s library partners, with some contracts indicating more egalitarian cooperation between the library and the company, and others explicitly granting Google the final say (e.g., University of Michigan and Google 2004, University of California and Google 2006, University of Texas and Google 2007).

4.1.2 Library Pre-processing

For the libraries, one of the most labor-intensive pieces of the scanning process seems to have been the retrieval and processing of books to send to Google – or, more precisely, performing this work quickly enough to “*feed the beast*,” to keep the scanning center adequately supplied with books. The exact nature of the pre-processing stage varied among institutions, but generally seems to have included performing basic conservation work to make sure the book could handle transit and scanning, cleaning up the book itself (i.e.

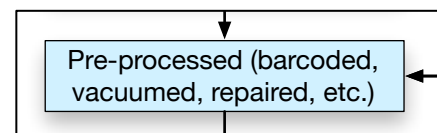


Figure 43

¹⁶⁸ This phrase is an actual quote from a participant (P16). Other similarly colorful expressions of the way in which the scale of the project dominated library processes include an OCA informant’s account of an NYPL librarian saying of GBS that, “*Every day all we do is we wake up to feed the firehose*” (P11), and a Harvard librarian’s description of it as “*the 900 pound gorilla project that chased everybody else out of the room*” (P13).

vacuuming it on all six sides) to help improve the quality of the scan, and ensuring that the book could be automatically tracked via a barcode or similar marking. Remarkably, the turnaround time on volumes – the time elapsed between their removal from the shelf and their replacement on it – could be as short as six days. As one librarian observes, this is rather “*amazing,*”

Considering they get pulled from the shelf, then they have to be vacuumed, and the barcodes scanned, and... and managed on a cart, and the cart has to get moved to a loading dock, and everything get queued up and it goes to Google, and Google had to check all that... check it all in, in its... system... and then get it queued up for the next days' work, and back out the other direction. And then we receive them back, had to get them, put them in, find, sort, get them on the shelf, so six days was mighty fast (P17).

Still, as remarkable as this turnaround time may have been, achieving it required a great deal of very fast work on the part of each library – and this proved more of a burden for some libraries than for others.

In particular, at Oxford and Harvard, and to some extent also NYPL, relatively large portions of the collections were deemed unready to leave the building, in terms of condition, trackability, or both – and this made the entire process a great deal more labor intensive (P13, P16, P18). As one Oxford librarian explains,

So they [Google] designed their scanning center with a certain workflow in mind, and being able to keep up with the scanning center was... was a real challenge given that our collections weren't like many of the US libraries [...] Ours were [...] spread over multiple different storage locations. Some of them dating back to the 17th century, many of them you know which involved very serious kind of manual handling issues. And so just keeping pace with the demands of the scanning center so that they had enough books and enough books in the buffer was a constant challenge and one which we wrestled with every single week for three years (P16).

And at Harvard,

lots of the material in the Widener collection are either serials or reference materials which have never been allowed to circulate. So they didn't have barcodes. You know, like, say 4 million volumes in the building... at least one and a half million had no way to be managed in an automated fashion. So we had to set up a project where we went through the Harv---the Widener Library stacks and barcoded things, and cleaned up bibliographic records, and you know... Basically engaged in a lot of inside work to actually make it possible for this outside project to function (P13).

Although this process of preparing materials for scanning was extremely challenging and often aggravating for the librarians involved, however, it actually turned out to have benefits beyond enabling Google's scanning project. For example, the same Oxford librarian cited above went on to note that the “*housekeeping intelligence*” gained as “*an unexpected byproduct*” of this process helped the library to actually locate several thousand 19th century books that had not been cataloged at the time of their acquisition, and that the library thus

had not previously known they owned (P16). And the librarian I spoke with from Harvard observed that at that institution, the challenges of participating in this project seemed to help in unifying the previously extremely disconnected libraries at Harvard – in “*turning the Harvard libraries into a Harvard library*” (P13).

Additionally, to some degree, the amount of work that went into pre-processing was likely linked to the relative conservatism of the institution regarding what sorts of materials, in what condition, the institution was willing to send to Google. As noted in the historical synopsis, there was an impression among the partners that Michigan was the most liberal, and Harvard the most conservative, with Stanford somewhere in between (P17). It is not entirely clear where Oxford and NYPL fall on this spectrum, but based on available evidence, I would place NYPL closer to Harvard, and Oxford closer to Stanford. Given a more conservative attitude (and/or mistrust of Google), it would make sense that more time might be spent preparing the books for the potential travails they might face in the scanning process – thus increasing the workload required to “*feed the beast*” for librarians at more conservative institutions.¹⁶⁹

4.1.3 Transit and Scanning

There are many lacunae in the data regarding the process of transporting and scanning books for inclusion in Google’s initiative, due mainly to the non-disclosure and general secrecy issues mentioned frequently throughout this chapter. This section will thus present what can be established, but will also underline the gaps that remain.

Generally speaking, in each Google partner library, books to be scanned by Google were loaded onto vehicles and transported to Google scanning centers in undisclosed locations, scanned (or, increasingly as the project went on, rejected) by Google employees, and then sent back to the library. As one Michigan librarian describes, although the elements of book scanning and transport may seem like they would have already been well worked out by libraries, Google found ways of tweaking particular pieces to further improve efficiency:

Google would come in, and take out tens of thousands of books, I can’t tell you how many, I signed an NDA, um... several tens of thousands of books, put ‘em on trucks, take them to their secret location, which you could find if you followed the truck in your car, um, and scan ‘em and they had, they had these, they designed – this is so Google-like – they designed their own book carts. You’d think the book cart would have been optimized, NOPE, the Google book cart is actually better for carrying around in a particular size of truck. So they... the book carts roll into their thing, the books come off in order, they get processed, they come back onto the book cart, the book cart goes back into the truck in the same, in reverse order of how it came out, you know this is sort of... this is how high school graduation works... um, and um, and it comes back to us. That’s much, much, much, much easier than working off of pick lists. And... the bulk (literally bulk) of the

¹⁶⁹ It is no coincidence, I think, that the three depictions of the GBLP offered in the previous footnote are linked, respectively, to Oxford, Harvard, and NYPL.

scanning that happened at Michigan was in that form (P10).

Although as he notes, this description may in some aspects only apply to the University of Michigan, it seems clear from both the contracts and other interviews that the process was not far off from this elsewhere – the main divergence would have been in other institutions’ selectivity at the point of loading the trucks, as discussed in the Collections section, above. Overall, this process seems to have been extremely standardized:¹⁷⁰ as a Googler involved in working out the library agreements explains,

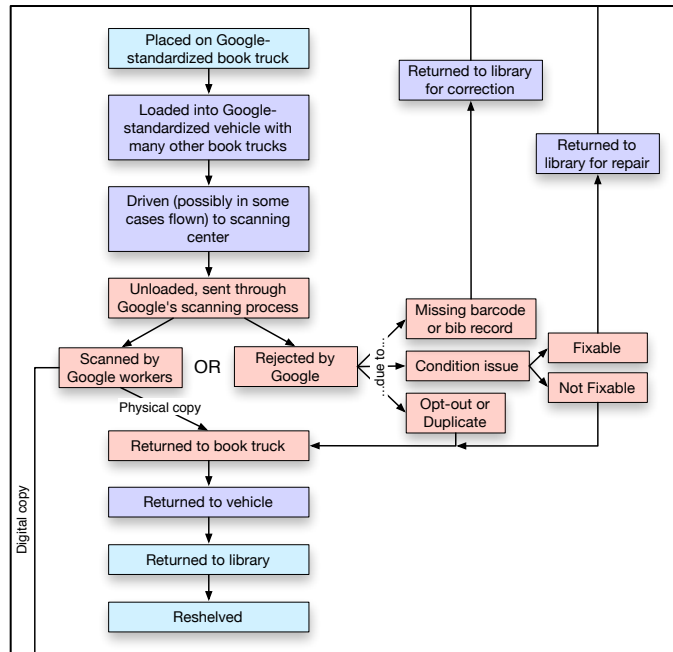


Figure 44

We agree early on on how long we're going to hold the books, what books we will scan, what books we won't scan. We really want to be careful that we don't do any damage to the books...and so we work closely with the libraries to sort define the limits of what condition is suitable for scanning and not suitable for scanning. We set up regular shipment schedules. We have expectations in terms of the number of volumes that we receive... um... you know, every period. Every... every interval. And that might vary from library to library, depending on their own capacity. But once it's in place we kind of keep it standard, and then the scanning methods that we use are standardized, um, so that wherever we scan a book whether it be in Europe or in the U.S., we're using the same procedures, the same protocols, um, and... you know, so the libraries understand very clearly exactly what's happening with their books. And when they're going to get their books back. So yeah I'd say it's quite standardized and... there's a pretty well-understood procedure now (P12).

Notably, moreover, this process appears to have been executed with remarkable care: although there are various references to damage liability throughout the public contracts, none of the librarians I spoke with reported any materials actually having been damaged. On the contrary, a librarian from Oxford, where Google was allowed to scan older and more fragile materials than they encountered elsewhere (P12), specifically noted that Google “didn’t lose any books, [and] didn’t damage any books in the process” (P16). If Oxford’s experiences are typical, then this would seem a fairly substantial accomplishment, given the scale of the project and the constant, circular flow of truckloads of books that it entailed.

¹⁷⁰ It was also, as noted earlier, remarkably efficient – a fact that participants credit both to input from the first five library partners and to Google’s early decision to hire a logistician, Doug Kuch, to streamline the workflows (P8, P12, P16, P17).

Within this process, there are two pieces that have remained especially obscure, and thus merit a bit more examination: (1) exactly where the books were being taken and (2) exactly how the books were being scanned.

4.1.3.1 *Scanning Location*

The public contracts give only a patchy indication of where the books are actually taken for scanning. All specify that it will be a secure facility, designated and run by Google (University of Michigan and Google 2004, University of California and Google 2006, University of Wisconsin and Google 2006, University of Texas and Google 2007); at least one, at the University of Wisconsin, indicates that the University will attempt to find space in a university building, and my informant at Oxford told me that University space was provided for Google there also (P16). All of the scanning locations are secret, even from most library staff,¹⁷¹ though one suspects they would not be too difficult to uncover, as P10 suggests above, simply by following one of the trucks.

It is not at all clear from available information either (a) how many scanning centers Google has utilized and/or constructed over the course of the project or (b) where Google scanned books from each individual library. However, a few insights from interviews can shed some dim light on at least some of the relevant facts. First, participants identified four specific locations where scanning centers were known to be: in order of establishment, Ann Arbor, Michigan; Mountain View, California; Cambridge, Massachusetts, and Oxford, England (P8, P12, P16, P17). Not coincidentally, each of these locations stands in close proximity to one of the first five Google library partners – Michigan, Stanford, Harvard, and Oxford, respectively. It seems safe to say that, at least at first, each of these institutions' books was scanned at the scanning center nearest to it (and potentially located on university property). It is not completely clear where books from the New York Public Library were taken to be scanned, though an interview conducted with an individual there suggests that it was not centrally located within the NYPL system (P18). It is possible (though this is pure conjecture on my part) that they were taken to Cambridge, about four hours away; however, it is equally possible that there was a fifth scanning center in or near New York, which simply did not come up in the interviews. As for later partners, I can only speculate: it could be that each one had a local scanning center of its own, but it seems at least equally possible that all institutions in a particular radius shipped materials to a central point – e.g. the CIC institutions to Ann Arbor, the University of California system to Mountain View, and so on. Still, however many scanning centers there might have been at the project's peak, another revelation from the interviews was that as of 2012, Google had only one scanning center still in operation in North America – the very first, in Ann Arbor. At least some of the

¹⁷¹ As I can attest from my time at the UM's Hatcher Graduate Library reference desk during the early years of the project.

project's U.S. participants – including Harvard and the NYPL – had finished all the scanning they had signed up to do with the company, which obviated the need to maintain scanning facilities near those institutions (P13, P18). However, this was not true of all. In addition to the University of Michigan, both Stanford and the University of California were still sending books for Google to scan as of October 2011, when the Mountain View scanning facility was shut down (P3, P17). After that point, and for as long as those institutions continue to scan with Google, those books must make a cross-country trek (by plane, train, or truck, I do not know) to Michigan for processing, and then another trek to get back (P17). Interestingly, although the pace of Google scanning overall has slowed considerably over the past few years (since the rejection of the settlement), its shelf-to-shelf turnaround times are still incredibly fast. Where books at Stanford had gone from shelf to scanner and back in six days when the facility was nearby, the duration of that trip had only increased to eleven or twelve days with the facility halfway across the country in 2012 (P17).

4.1.3.2 *Scanning Technology*

Discussions of the secrecy surrounding Google's scanning technologies in the interviews were often accompanied by expressions of exasperation (eye-rolling, sighs), by both librarians and Googlers. On the one hand, the proprietary technology that Google had developed to scan the books was one of the most closely-held secrets of the entire project (right alongside any and all numerical data). The most that the company or its partners would ever say about the technology or the scanning centers was that nobody was allowed to see either. For example, the UM FAQ about the project includes the following Q and A:

Q. 15: Can we see the scanning facility?

A: Google does not permit access to the scanning facility. Because of the proprietary hardware and software in use, Google prefers that the processes and equipment not be made public (2005).

And extremely similar language appears in the FAQ for the CIC partnership (2008).

On the other hand, however, the rationale for this high level of secrecy was never entirely clear. As the discussion of Google's scanning patents below will show, although the company's scanning technologies were undoubtedly clever, they were hardly earth-shattering within the field of book digitization – and moreover, given the relatively low return on investment for book scanning, it is far from clear that Google would ever have faced significant competition in this area. As such, many both involved in the project and observing it from within the surrounding library community found Google's unwillingness to share on this point fairly perplexing. One UM librarian, when asked about the company's secrecy on numbers and scanners, responded, "*Oh yeah, the scanning technologies and the numbers. God!*" and went on to elaborate,

What I wound up saying to people about the scanning technology, and Google never got mad at me for this, was... if you saw it, you'd say, "that's really smart." But you wouldn't say "oh my God, how did they think of that." Um... and that actually... and if you can't figure out, knowing a little bit about how this works, from that statement, then too bad. So I would've... Clearly they didn't want other people to have um, y'know, advantage over them, I think that there's some dim sense of corporate intelligence which you don't want to tell people how much you're investing in something... ahh... and so they had those concerns. I didn't think that any of those were really, really, material. They were just annoying (P10).

And interestingly, one of the Googlers on the project concurred that the level of secrecy surrounding the scanning technologies was likely unnecessary, and even suggested that, in retrospect, more transparency might have helped the project's public image:

The logic of not being transparent was...well our competitors will know what we're doing and they might start doing the same thing. I never really thought our competitors would be interested in doing the same thing [...] and they never did, and I don't think it's because...y'know, that if they would've had more exposure to our technology, or our plans, that they would've been behaved differently. So I guess in hin- y'know I always was an advocate for more transparency, but to be honest, I don't think knowing more about the technology that we were using for scanning was gonna have a huge difference, but I think it might- the perception of lack of transparency on important things? And I think that was negative (P4).

Of course, two competitors – Microsoft and Yahoo! – *did* in fact take a run at book digitization, as supporters of the Open Content Alliance. Yet, both of these companies also abandoned their efforts quite precipitously after only a few years and a few million dollars (or, in Yahoo!'s case, a few thousand) – a fact which lends credence to this participant's assertion that sharing the details of Google's scanning technologies was unlikely to spur greater competition on the part of these rival entities.

Whether or not Google's initial secrecy was justified, however, it had become largely irrelevant by the time of this writing: thanks to the disclosure requirements of the patent process, Google's scanning technologies have actually not been much of a secret since 2009. Google initially filed for a patent on its scanning system in September of 2004 – before the library project was even announced – and five years later, when the patent was granted, the system became a matter of public record.¹⁷² The basic components of this system, according to the patent, are:

- an infrared projector;
- a pattern mask located in an output path of the infrared projector;
- a stereoscopic camera to generate images of opposing pages of a document from which a three-dimensional image of a surface of the opposing pages of the

¹⁷² The filings are posted online through both the USPTO and Google itself (by searching patent number 7508978 at <http://patft.uspto.gov/netahtml/PTO/srchnum.htm> or directly at <http://www.google.com/patents/US7508978>)

document is generated; and
control logic to process the three-dimensional image to locate a groove that is defined by the spine of the document and located between the opposing pages of the document (Lefever and Saric 2009, col. 6).

Essentially, the system uses multiple cameras and three-dimensional modeling to find the spine of the book and un-warp the natural bend of the book pages as the book lies flat on a table. Where a normal photo of an open book, taken from above, will show the edges of the pages as curved lines, this method re-squares the pages, which significantly eases the character-recognition process.¹⁷³ Figure 45 depicts, in basic terms,

the positioning of the book, platform, sensor, projector, and cameras, in which the rectangles labeled “305” and “310” are high definition cameras. The patent also notes that in some instances only one high definition camera might be used, and in others, more than two; several other minor variations like these are also included in the patent (Lefever and Saric 2009, col. 2-3).

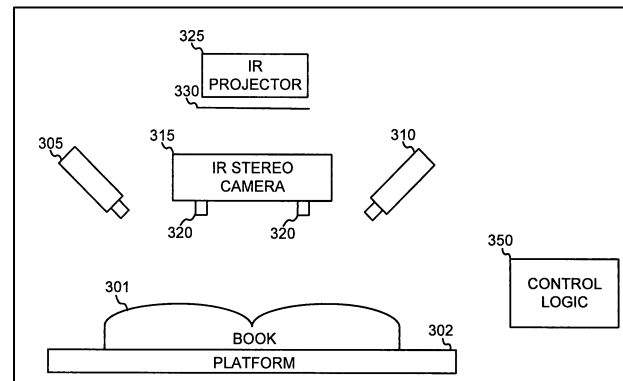


Figure 45: Depiction of Google's Book-Scanning Technology from U.S. Patent #7508978

This sort of technology would logically help to make the scanning go more quickly, as it offloads much of the work of straightening the book and holding the pages flat onto the software: for example, if the book is held a bit crooked, the software can still find the groove in the book, and work from there to figure out how to deskew the lines of text. Also, if the cameras involved work similarly to a typical DSLR – that is, they are able to take each shot in a fraction of a second – that would be also a significant gain in efficiency over methods that involve running a scan head more slowly across the surface, as in a typical flatbed scanner or copy machine.¹⁷⁴ It is also, notably, a very “Googley” way of doing things: making as many pieces of the process as possible operate algorithmically and automatically, without the need for human intervention.

¹⁷³ I can attest to the value of this type of approach, based on my own experiences gathering archival materials for the Boston and Carnegie cases described in earlier chapters. Many, if not most, of the sources I worked from were photographs of open books – except rather than being taken with a fancy stereoscopic camera on a fixed arm, they were mainly taken with a Canon DSLR, held by me. The post-processing required on these images to make them OCR-able (when they were even in print, and not handwritten) was fairly arduous, and required post-engineering something very much like the deskewing process described here, but in a much more manual, piece-by-piece way.

¹⁷⁴ Indeed, at the CCNY archives at Columbia, this was one of the primary reasons I resorted to photographing the microfilm reader screen to capture the content I needed, rather than using the microfilm scanner – I could click a photo almost instantaneously, while the scanner would take ages to go over each page.

Somewhat less Googley are the elements of this process that *did* require human intervention, and in particular, the employment of likely hundreds of contract workers to turn the pages of the books. As Levy explains,

Google found that the state of robotics did not allow for a speedy process by which a machine could turn the pages itself without shredding them. So despite the fact that hiring a wave of human laborers did not conform to Google's scaling philosophy, humans it was. Every so often, one literally would see the fingerprints of the Google worker in charge of the task on the scans (2011, 350-51).

One of the few requirements for the scanning process, both specified in the available library contracts and noted in library statements, was that the process had to be non-destructive with regard to the books (University of Michigan and Google 2004, Stanford University Libraries 2006, University of California and Google 2006, University of Wisconsin and Google 2006, Universiteitsbibliotheek Gent 2007, University of Texas and Google 2007, University of Texas Libraries n.d.) – or at least, no riskier than lending the books out to undergraduates. As the University of Texas FAQ asserts,

Google has made an extraordinary commitment of research and technology to ensure the books are not damaged. We have confidence in Google's expertise and experience in handling physical and digital texts. Any time a patron or librarian handles a book, there is a risk of damaging it. That's a reality of offering access to collections. By digitizing the books now, we can reduce further damage from human handling, mold, insects, etc. (University of Texas Libraries n.d.).

Employing human workers to turn pages was one of the ways in which Google upheld this commitment. However, this piece of the project also attracted a fair amount of criticism, related to the working conditions of the actual people doing the page-turning (e.g., Wilson 2009-2011, Barrett 2011). Though some of these critiques have been disputed since they first emerged in 2011, the concerns over the class system within Google remain significant, and thus worthy of mention – though not extensive discussion – here.

4.1.4 Digital Delivery

Although the physical volumes are returned to the libraries within a matter of days, delivery of the digital version of the book can take much longer. Within the public contracts, several different delivery windows are specified, depending on the institution. Two,

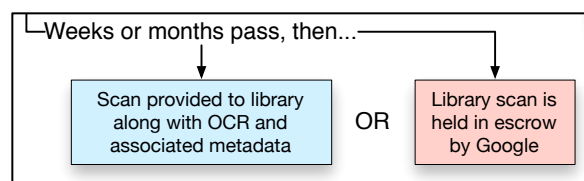


Figure 46

with the University of Wisconsin and the UVA,¹⁷⁵ do not specify any particular timeline, while the CIC contract indicates that scans should be delivered in a time “not to exceed six months from date of scanning” (University of Wisconsin and Google, Committee on Institutional Cooperation and Google 2007, University of Virginia and Google 2010). The Michigan, Texas, and California contracts all seem to indicate a narrower window, specifying that the digital versions should be delivered to the libraries within 30 days after the content has been “Successfully Processed” – a term that is taken to mean, “when Google determines it has satisfactorily gone through all stages of Google’s digitization, post processing and quality assurance procedures” (University of Michigan and Google 2004, University of California and Google 2006, University of Texas and Google 2007). However, the amount of time required for this “successful processing” to be completed is considerably fuzzier, and may be considerable: in fact, a follow-up conversation with an interviewee from UM revealed that the time elapsed between the scanning of the book and the delivery of the digital files can sometimes be several months, with some volumes remaining in unexplained “Google limbo” for significantly longer than others scanned at roughly the same time (P10).

Whatever the timeline, however, all of Google’s library partners did contract to receive digital copies of the items Google scanned from their collections (P12), and most, if not all, also negotiated the right to share those digital copies through consortial arrangements with other libraries (e.g., University of Michigan and Google 2004, University of Wisconsin and Google 2006, Committee on Institutional Cooperation and Google 2007, Giersberg 2007). For institutions only allowing the scanning of public domain works, these transfers were generally unproblematic: in the absence of copyright, a work can be copied and re-copied as much as anyone may like. However, for institutions allowing Google to scan in-copyright works, an additional path for those particular digital files emerged. That is, Google offered to hold digital versions of in-copyright works in escrow until the legal issues surrounding the scans diminish to the point where the risk to the universities in possessing those files becomes acceptable.

It is unclear how many institutions took Google up on this offer, but my interviews as well as other sources establish that Stanford and the CIC were among them, while the University of Michigan was not (P10, P17, Committee on Institutional Cooperation 2008). The escrow process for the CIC is laid out in their Cooperative Agreement, and allows for release of items held in escrow to the source universities under five conditions:

- (a) the In-Copyright Work becomes in the public domain;

¹⁷⁵ This is the amended Virginia contract, released around the time of the settlement; the original no longer appears to be available on their website. Given that the purpose of the amended agreements was to incorporate the provisions of the settlement agreement, and that the timeline of digital delivery does not have anything to do with those provisions, it seems likely that this section of the contract remained the same from one version to the next.

- (b) a Party has obtained permission through contractual agreements with copyright holders that includes the right to make a copy of the In-Copyright Work and to provide it to the CIC or Source CIC University;
- (c) well established case law exists that In-Copyright Works can be copied and held by the CIC Administrative Offices and/or the Source CIC University without infringing on the rights of a copyright holder;
- (d) if at any time Google is in material breach of its obligations under Section 4.3(b) or 4.6(a) and Google does not remedy any such failure within ninety (90) days after its occurrence (or, in the event such failure is caused by technical problems or causes similar to those described in Section 12.5, within such longer period as Google, working diligently, reasonably requires to remedy such problems); or
- (e) the CIC Administrative Offices or the Source CIC University and Google agree in writing that the release of a particular In-Copyright Work or Works is legally supported and appropriate under the terms of this Agreement (Committee on Institutional Cooperation and Google 2007, section 4).

While the reduction in legal risk offered by the escrow process has certainly been appreciated by some universities, one librarian I spoke with also expressed regret that under this system, her library could not have direct access to the scans made from its collection, at least for the time being:

P17: Certainly having the books indexed on the inside is great. And as time goes on I hope we will have the opportunity and the legal right to download these files. Content files. For preservation.

EAJ: Yeah, you had mentioned something about Google holding the files in escrow.

P17: Yeah, for which of course we're very grateful. I wish we could um... I wish our counsel would let us. But. I don't want to be sued, either, so I'm... as I say I'm a very understanding person. But as I said, if one of the major motivators was to get those files then not having them makes me sad. That's not to say we'll never have them. At all. God help me. I couldn't live if that were the case. But I would... it would be nice if we could come to some legal position where we felt we could download them all.

At the time of this writing, the Author's Guild's lawsuit had only recently been decided in favor of Google, finding the scanning and the library arrangements to have fallen under fair use. It is possible that this decision may open the door for the release of these files from escrow; however, as the case remains under appeal, it is unclear whether or when that might ultimately happen.

4.1.5 Constraints on use

The final piece of the process of getting a Google book from shelf to user is to actually make the digital version accessible online, through Google or library websites or both. The particular forms taken by this online access will be described further in the discussion of structures, below; however, there is an element of process here as well. That is, in constructing the library scanning agreements (and later the interfaces themselves), certain constraints on use were developed, agreed upon, and baked in to the access systems

through which the books were provided. These constraints fall into two basic categories: (1) constraints on the libraries' use of their digital copies (and through them, on library users' uses of those copies), and (2) constraints on end use of the Google digital copy.

4.1.5.1 Constraints on library use of library copy

The restrictions on library use of library digital copies stem from two basic edicts: don't make money off the scans, and don't make it easy for anyone else to make money off the scans. In practice, these conditions required that the libraries close the door to competing search robots, refrain from sharing the digital copies too broadly with other libraries or the public, and obstruct automated/bulk downloading of the files. For example, the University of Michigan Cooperative Agreement stipulates that:

U of M shall implement technological measures (e.g., through use of the robots.txt protocol) to restrict automated access to any portion of the U of M Digital Copy or the portions of the U of M website on which any portion of the U of M Digital Copy is available. U of M shall also make reasonable efforts (including but not limited to restrictions placed in Terms of Use for the U of M website) to prevent third parties from (a) downloading or otherwise obtaining any portion of the U of M Digital Copy for commercial purposes, (b) redistributing any portions of the U of M Digital Copy, or (c) automated and systematic downloading from its website image files from the U of M Digital Copy. U of M shall restrict access to the U of M Digital Copy to those persons having a need to access such materials and shall also cooperate in good faith with Google to mutually develop methods and systems for ensuring that the substantial portions of the U of M Digital Copy are not downloaded from the services offered on U of M's website or otherwise disseminated to the public at large (University of Michigan and Google 2004, section 4.4.1).

Very similar language appears in the other public contracts. Under these terms, there are some restrictions on what the libraries can do even with materials that are in the public domain, and thus free of other restrictions based on copyright. Librarians' perceptions of the onerousness of these restrictions, however, vary considerably. One, at UM, cited the limitations on the use of public domain materials as his primary regret, regarding both the negotiation of the initial scanning contracts and the failure of the AAP/AG settlement agreement:

The biggest thing in the project that I wish had happened differently was the, um... the ah, agreements on the public domain. So... The agreement on the public domain that we're under now, um, is we can... post it, we can put it up, we can make it available on a reader, um... we can allow limited download, but we can't allow bulk download. Um... and indeed the rule, the deal between us and Google is something like, um... it can't be made available to potential competitors, um, the way, it you now... so we have to take efforts to make it slow to copy. Is the essential piece. And there's no expiration date on that. [...] I'm perfectly comfortable with deals where we trade off a little bit of now for a lot of forever. Um... If I had been directly involved in the contract negotiations with Google at the time on just this question I might have stuck my... dug my heels in, um,

but I was actually off doing other things. I wasn't involved with the library contracts. As part of the settlement agreement, we actually did put a time limit on the public domain restrictions. Uh, it wasn't great, but at least there was a limit. But now the settlement agreement hasn't gone through and won't. So... we don't have that anymore. We're still at infinity, which is a little long. But that's I think my biggest regret of the way it played out. Was that we didn't at least put some, you know... hell 2050! 2050 would be stupid, but 2050 is sooner than never (P10).

On the other hand, however, a librarian at the University of California expresses a very different opinion, emphasizing the extent to which he believes that the rhetoric of Google “locking up the public domain” has been overblown by critics of the project and/or the settlement:

I feel, um, the Google agreement, I feel very proud that we got them to release the public domain. Very---um, vir- um, frequently, and incorrectly, and intentionally so, represented by Brewster and Pam Samuelson, other detractors of that great entity. [...] you know, they claim that they had put a lock on the public d---which was entirely untrue---and it was entirely untrue because the University of California agreement not only gave us complete freedom to do anything we wanted with the public domain, uh, but Google then, immediately after signing it---and we knew they were going to---agreed to make the public domain works freely available through Google Book search to anyone, forever, for free. That was pretty good (P3).

Notably, the University of California contract contains virtually identical provisions to the University of Michigan's regarding restrictions on bulk downloading and commercial exploitation, and California's consortial sharing restrictions seem, if anything, stricter than at Michigan (University of Michigan and Google 2004, sections 4.4.1-4.4.2, University of Wisconsin and Google 2006, sections 4.9-4.10). Still, to an extent, my impression is that both of these individuals have a point. On the one hand, it is regrettable that there are *any* restrictions on the public domain, regardless of what those restrictions are or whether they are actually likely to impact the average reader's use of public domain books. On the other hand, the criticisms lobbed at the project for its restrictiveness on the public domain *do* seem hyperbolic. In my own interview with him, for example, Brewster Kahle asserted (not for the first time) that the GBLP was a “lock down commercialize the library system project,” and accused Google, Harvard, Stanford, and the University of Michigan of creating a system to “lock up the public domain” and establish “a walled, higher education, licensed, screwed-down library.”¹⁷⁶ It seems to me that there is a fair distance between the *actual* restrictions placed on the public domain within the project – centered mainly on preventing secondary commercialization – and the *rhetorical image* of those constraints, which casts them in a much harsher – and arguably inaccurate – light.

¹⁷⁶ Brewster Kahle waived all participant confidentiality; he is alone in doing so among my participants.

A final piece worth mentioning here is that although the University of Michigan contract does not include any fixed sunset date for the restrictions on its use the library digital copies, it does include a termination clause. That is, if Google ever stops providing free search of the digitized works – for example, if Google decided to shut down Google Books, as it has so many other services – that action would terminate all restrictions on UM’s use of its copy (University of Michigan and Google 2004, sections 4.3-4.4). Interestingly, however, UM seems to be the only university to have negotiated this particular escape-hatch clause, at least among those with public contracts: there is no similar clause in the contracts for UC, UT, or UW¹⁷⁷ (University of Wisconsin and Google 2006, University of Texas and Google 2007). And while the CIC contract specifies that Google must provide free public searching as long as Google is doing anything with the scans (Committee on Institutional Cooperation and Google 2007, 4.3b), it does not specifically release CIC institutions from the constraints on their own use of the scans if Google fails to provide such access. If Google shuts down Google Books, then, it is unclear what would become of the restrictions on use for many of its library partners. It is entirely possible that were this to happen, Google would follow Microsoft’s example in releasing all claim to its scans; however, in the absence of assurances to this effect across several contracts, that decision would be Google’s alone to make.

4.1.5.2 *Constraints on end-use of Google copy*

Most of the constraints on end-use of the Google digital copy are enumerated in the Google Books Terms of Service (TOS), and I will not unduly belabor them here. The two major types of limitations outlined in the terms of service are (1) technological and (2) rule-based. Briefly:

- Technological limitations: Users need to have access to a compatible device, compatible software, and a functioning Internet connection in order to use the books; however, access is not restricted by device, at least not as a matter of rule (only as a matter of technological pragmatics) (Google 2011b).
- Rule-based/Legal limitations: These are a bit ambivalent. On the one hand, the TOS states that “You may not sell, rent, lease, distribute, broadcast, transfer, or assign your rights to the Digital Content or any portion of it to any third party except as expressly permitted by Google” – but in the very next breath, it adds the caveat that “nothing in the Terms of Service shall prohibit any uses of Digital Content that would otherwise be permitted under the United States Copyright Act” (Google 2011b). In the case of materials that are in the public domain, it seems as though none of the prohibitions in the first sentence would actually apply, since they are all uses that would be permitted under the U.S. Copyright Act for public domain works (which are free of restrictions on copying, distribution, sale, etc.).

¹⁷⁷ In fact, in some of these, the library restrictions survive even termination of the contract due to breach.

Beyond these limitations, Google does also claim a fairly broad license to user-generated content applied to its books (e.g. annotations, user-submitted metadata), and there is a chance that users may *feel* constrained in their reading behavior by the fact that all their actions on the service – every click, every page turn, every highlighted passage – are being closely watched and analyzed by Google – an issue which I have explored elsewhere at length (Jones and Janes 2010). And finally, it is worth noting that Google does include a list of “Usage guidelines” on the first page of every PDF-downloadable eBook it offers, which requests that users:

- “Make non-commercial use of the files,”
- “Refrain from automated querying,”
- “Maintain attribution,” and
- “Keep it legal”¹⁷⁸

And although many critics have asserted that this is a clear example of Google placing egregious limits on use of the public domain (e.g., Lenssen 2007, B. McCoy 2007), it seems clear from context (as well as other statements by Google¹⁷⁹) that these guidelines are merely *requests*, rather than actual, enforceable conditions of use. After all, just above this list, the company states quite plainly that “Public domain books belong to the public and we are merely their custodians.” Although Google has blocked automated downloading of the books (and has required its library partners to do so also), the company seems to acknowledge that it has no actual standing to prevent users from providing such services themselves – it can only ask that they refrain.¹⁸⁰

Additionally, there are at least two other limitations on end-user access through Google that bear mention. First, for several years, Google’s public domain eBooks were only offered for download in PDF format – but those downloads did not come bundled with the associated OCR. Indeed, although Google does now offer downloads for some books in ePub format,

¹⁷⁸ The particular cover page I have referenced here appeared on an 1897 edition of *Alice’s Adventures in Wonderland* (accessible at: <http://www.google.com/books?id=XlsVAQAIAAJ>); however, to my knowledge, all of these cover pages are identical.

¹⁷⁹ In fact, in response to a flare-up of criticism on this point in 2007, Google released a statement to *Search Engine Land* that read:

We have gotten this question in the past. The front matter of our PDF books is not a EULA [end user license agreement]. We make some requests, but we are not trying to legally bind users to those requests. We’ve spent (and will continue to spend) a lot of time and money on Book Search, and we hope users will respect that effort and not use these files in ways that make it harder for us to justify that expense (for example, by setting up the ACME Public Domain PDF Download service that charges users a buck a book and includes malware in the download). Rather than using the front matter to convey legal restrictions, we are attempting to use it to convey what we hope to be the proper netiquette for the use of these files (Sullivan 2007).

¹⁸⁰ And indeed, the large number of Google Books that have been uploaded to the Internet Archive by the Pirate Bay (IA user tpb) tend to confirm the extent to which Google does not and perhaps cannot exert actual control over where these PDFs go and how they are used (even if it wished to, which is questionable).

complete with OCR, this is not true for all – many still remain available only in PDF. For those books, in order to search the book, the user either has to return to Google (which requires connectivity) or OCR the book herself (which requires software capable of doing that). Though one can easily imagine why this limitation was put in place – forcing users to return to the site is a very good thing from Google’s perspective, after all – it does place substantial limitations on the utility of Google’s eBooks for users in settings with non-continuous connectivity and/or lacking access to fairly sophisticated PDF-editing software. Such users can still *read* the books, of course, but the functionality of the downloaded version is, for them, significantly degraded; this essentially creates a second-class user experience for readers who are likely already at an informational disadvantage, given the limitations on their access to technologies just noted.¹⁸¹ And second, various observers have noted – at varying levels of formality – that Google’s copyright analyses often seem overly conservative, keeping books out of public view that should rightfully be within it. For example, U.S. government documents are generally public domain by law, regardless of publication date;¹⁸² yet, as of the time of this writing, Google was still treating many (if not all) post-1923 government documents as if they were copyrighted works – that is, showing only snippets, and not the full text (e.g., Prelinger 2006, Townsend 2007). And for international works, as previously noted, Google often places the cutoff date for books to be presumptively within the public domain (and thus accessible in full text) a fair bit earlier than would be prescribed by conservative local estimates – in the UK, for example, Google places the bar at 1863, while the Oxford libraries would place it at 1888 (P16) – and has reportedly also blocked all non-U.S. access to many works that are clearly in the public domain worldwide, such as books published prior to 1700, in addition to imposing a hard cutoff of 140 years (i.e. 1863) for release of materials internationally (e.g., Graf 2006, Sands 2006, Suber 2006, Clark 2013). While it is easy to understand why Google may wish to err far on the safe side with regard to the actual release of full text books¹⁸³ – and while in some cases (as at Oxford), these books *are* being made digitally available internationally through

¹⁸¹ Notably, although the libraries receive some associated metadata along with their digital copies of the books, most do not receive Google’s OCR either. The only exception to this among the institutions with public contracts is the University of California, where the image coordinates are explicitly included in the library digital copy (though UC is not allowed to share those coordinates in any way, shape, or form, with anyone) (University of California and Google 2006, sections 4.7 and 4.10a).

¹⁸² There are some complexities here with regard to documents like congressional hearing transcripts, which may include works that actually *are* copyrighted within them; however, Prelinger (2006) found several examples of documents that would not seem to be subject to these complexities, and as of 2013, those examples are all still in snippet view.

¹⁸³ There may be a fair use or fair dealing argument for *scanning* the books in various countries, but making the full text of an in-copyright work publicly available online, even if done in good faith, would leave Google with no such defense. Additionally, once such a work was released online, there would be little chance of taking it back: the nature of the Web, after all, is to make innumerable copies, and pass them around forever. Regardless of how dearly one might wish to remove something

Google's partner libraries – it seems that at times Google errs beyond the point of reason, and thus unnecessarily hinders access to works that rightfully belong in public view, for public use.

4.2 Structures

The relevant structures for a digital information service – or perhaps more accurately, a set of digital information services – are not entirely analogous to those required for a physical institution. Physical libraries, as discussed in the previous chapters, have physical facades, rooms dedicated to particular purposes, shelving systems, and catalogues to help patrons find things; they are situated in specific, unmoving physical locations, and they actually, literally contain both the patrons and the staff of the institution, many of whom can observe one another's presence and activities. Despite serving several analogous functions, digital "library" structure is quite different. In place of a single physical façade, there is a potentially infinite variety of virtual entry points. No rooms exist to gather together people and activities of particular sorts; instead, a wide range of pages serve an equally wide range of purposes, each tailored to disaggregated individuals and their particular searching behavior. Each book – now digital rather than paper – may have a fixed position in a server's memory somewhere, but that position is both obscured from and generally irrelevant to the end user, since each book can be served up to her on her home computer, and simultaneously to millions of other users on theirs, even as it continues to remain lodged in the file system of that hidden server. There is a sense in which digital books need never leave their "shelves" in order to be read. And of course the catalogue can contain vastly more information – not just basic bibliographic data, but the full texts of the books themselves – and yet along with that expansion comes procedural complexities regarding accuracy, relevance rankings, and determinations of how much of that content can be revealed to users.

Given the constant mutability and extensibility of the digital structures that make up Google Books and other initiatives like it, any attempt to comprehensively describe those structures is bound to be futile; they are effectively infinite. As such, this final section of the chapter will limit itself to describing four pieces of that shifting web, proceeding roughly from the outside in: (1) the Google Books interface, (2) library services for finding and hosting Google-scanned books, (3) standards for quality and metadata, and (4) what I'll call "invisible structures" – the pieces of the project that remain obscured from public view, and likely always will.

4.2.1 Viewing Books (or Parts Thereof) on Google

Google offers four different "views" of its eBooks at the Google Books site: Full View, Limited Preview, Snippet View, and No Preview. The application of any of these views to a particular book depends on two factors: first, the copyright status of the book, and second,

whether it has been incorporated into the Partner Program¹⁸⁴ (Google 2011c). Each of the views, and conditions under which each is applied, are outlined below.

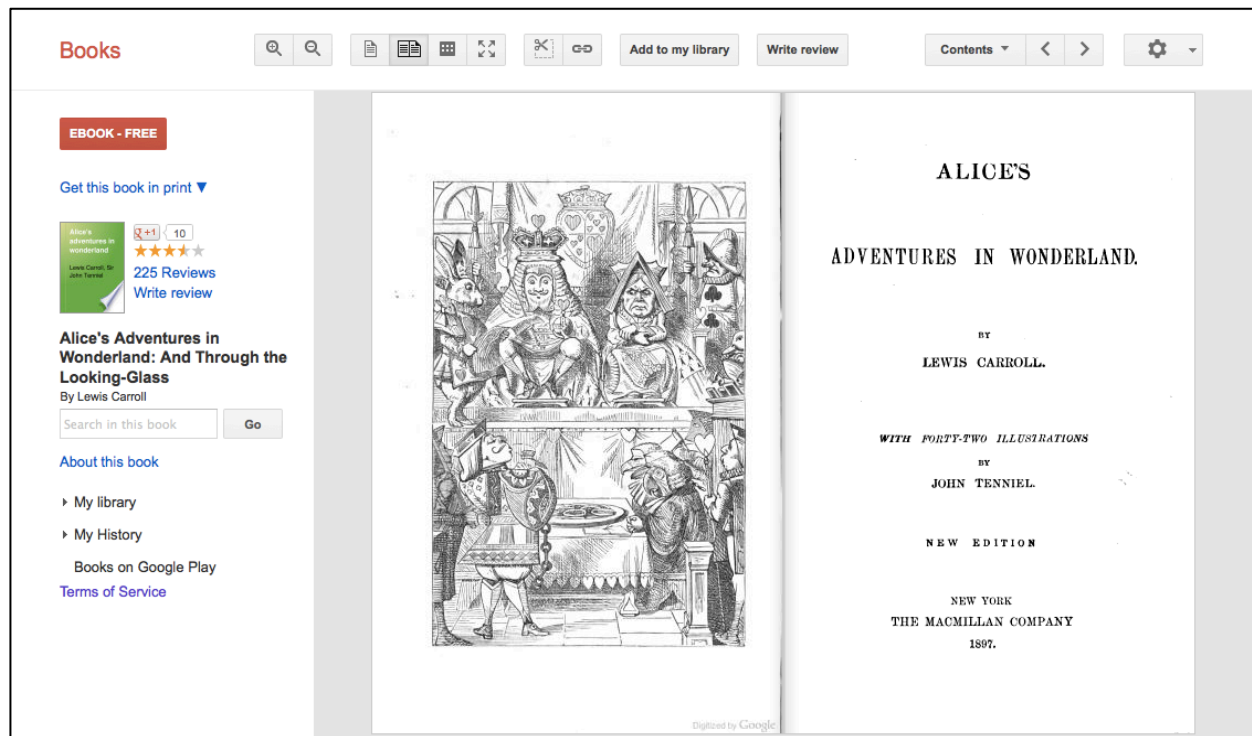


Figure 47: *Alice's Adventures in Wonderland* (1897) in Full View on Google Books (screenshot by author, July 2013)

In Full View (Figure 47), users are offered the opportunity to scroll through and read the entire book, and also to search within it; they also have the option of downloading the book in PDF, or in some cases ePub.¹⁸⁵ Full view is reserved for books from the library project that have been determined by Google to be in the public domain (as noted, often a very conservative judgment), and also for books that the copyright holder has explicitly given permission for Google to make available in full (Google 2011c).

In Limited Preview, a larger portion of the book than would typically be allowed by fair use

¹⁸⁴ Books can be incorporated into the Partner Program in two ways: first, they can be contributed digitally or physically by the publishers themselves, or second, they can be scanned from a library and then opted in to the Partner Program by their rights holder (Google 2012a).

¹⁸⁵ As I was writing this chapter, a few sample searches on Google Books revealed yet a fifth view designation, "Read," which seems, as far as I can tell, to mean what "Full View" used to mean. "Full View," by contrast, now seems to indicate a book that can be read in full online, but cannot be downloaded for free (some cannot be downloaded at all, others can be downloaded for a price). However, I was unable to find any discussion of this new distinction anywhere; indeed, the official Google informational page about the view types (Google 2011c) remains unchanged, and lists those four views as the only four. As such, I am sticking with the old designations for the purpose of this writing – this simply serves as a caution that most of the technological features discussed in this chapter are and always have been subject to change at any time. (It's also possible that I happened to catch the "B" side of one of Google's famed A/B tests, and that this is not actually a permanent feature at all.)

is offered (usually about 20%, but sometimes more or less), by agreement with the copyright holder. These digital versions are often – though not always (Google 2012a) – contributed by the publisher, and tend to be the nicest-looking images on offer through Google Books – likely because publishers can often provide born-digital versions, rather than scans. Books in limited preview are often offered for sale through Google’s eBookstore, and always contain a “Get this book in print” dropdown alongside the text, with links to both retailers and library search services (Google 2011c).

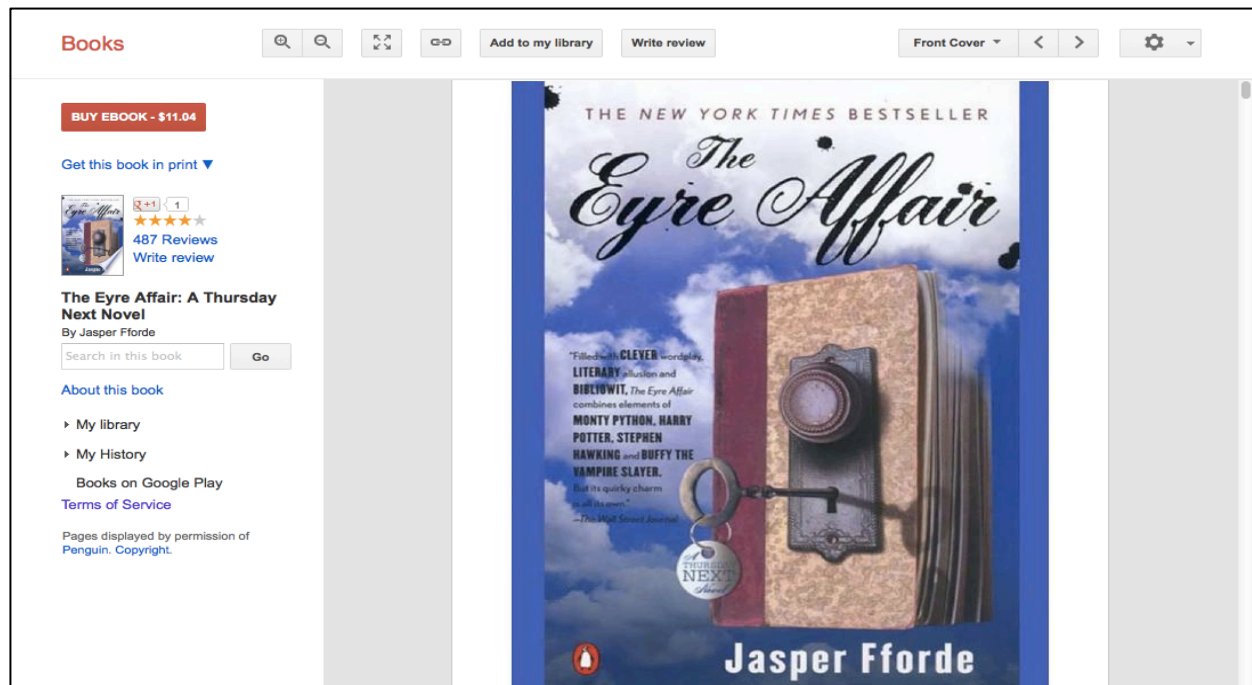


Figure 48: *The Eyre Affair* (2003), in Limited Preview on Google Books (screenshot by author, July 2013)

The third interface, “Snippet View,” shows no more than three sentences from a book, providing only the immediate context surrounding a search term highlighted in the center of the snippet. This view is used for books scanned as part of the library project that are (or may be) in copyright and have *not* been opted out by their copyright holder – a category which includes all orphan works whose public domain status has not been established. Orphan works, after all, are works whose copyright ownership status is unclear, and if the ownership is unclear, there is nobody in a position to either opt the book in to the Partner Program or opt it out of the Library Program. And so Snippet View persists in this niche – at least for now. During the interviews, more than one participant noted that snippet view had been deemed a poor user experience, and that as such, Google was losing interest in supporting it (P8, P9). Although it was unclear from these conversations exactly what Google’s proposed solution to this issue might be, it seems at least plausible that snippet view may be either disappearing or getting a significant makeover some time in the near future. For the time being however, it looks something like the image in Figure 49.

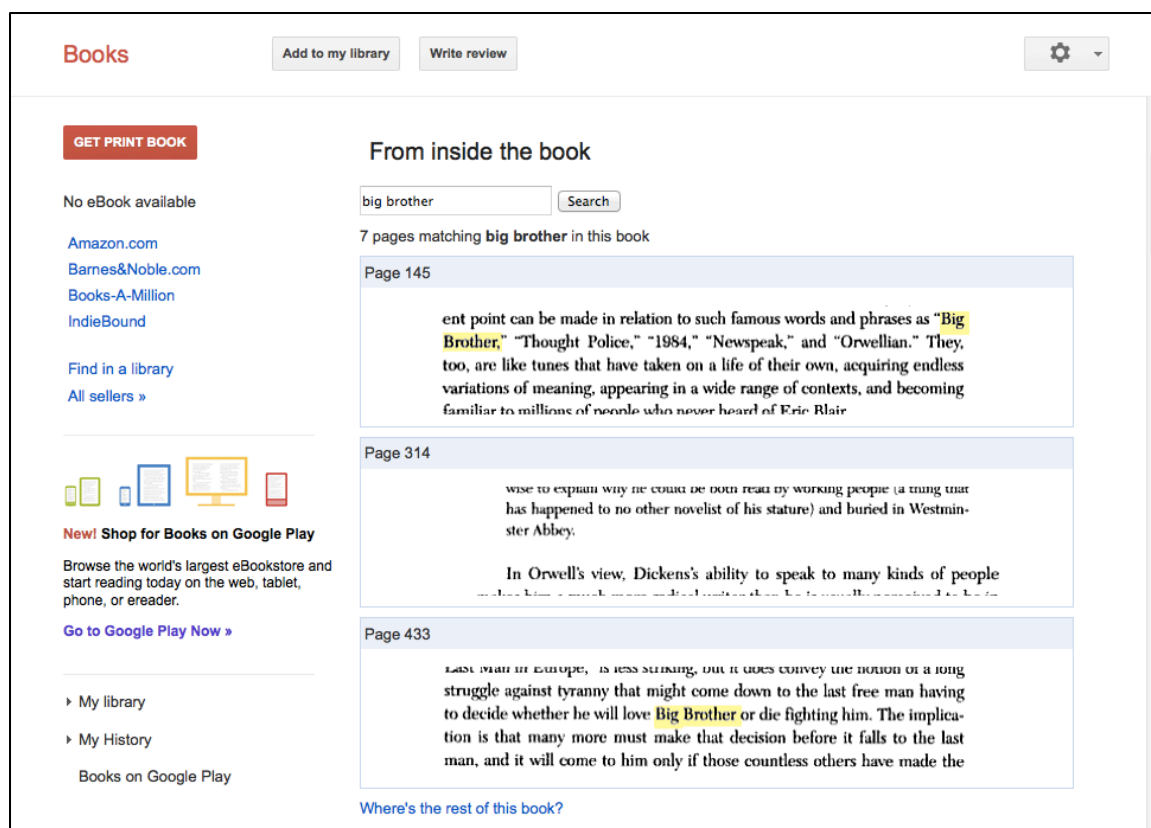


Figure 49: A search for “big brother” in *Orwell: The Authorized Biography* (1991), in Snippet View on Google Books (screenshot by author, July 2013)

Finally, Google Books also contains a large number of book records with no preview attached. For these books, Google shows only the metadata, in much the same way book information might appear on an Amazon or Goodreads page.¹⁸⁶ Notably, this category of view may well include books that were opted out of scanning. Because metadata – in this case, basic facts about a book’s title, authorship, etc. – is not generally copyrightable, authors and publishers have little say over what anyone might do with metadata describing their books.¹⁸⁷ Indeed, Google explains as much in its “Help” article on how to remove a book from Google Books, noting that if a book is labeled “No Preview Available,”

this result does not contain copyrighted pages from the book itself — instead the record shows information about the book, including licensed bibliographic data such as the ISBN, author, publisher, and publication date, as well as search

¹⁸⁶ For example, similar data for book shown in Figure 50, *Harry Potter and the Sorcerer’s Stone*, is provided on Amazon at www.amazon.com/Harry-Potter-Sorcerers-Stone-Book/dp/0590353403, and on Goodreads at http://www.goodreads.com/book/show/3.Harry_Potter_and_the_Sorcerer_s_Stone. (However, unlike Google, Amazon was able to secure the rights to allow users to “Search inside the book” for this volume. And it’s also worth noting that the reviews shown on Google actually originate from Goodreads, at least in this case.)

¹⁸⁷ It would be a woeful thing if such a right did exist. Consider: book authors could have prevented the construction of the BPL’s card catalogue, or any other kind of book-finding aid. Where would libraries have been then?

results from public websites. In effect, this is like seeing a library card catalog online. ... In general, we do not remove metadata-only results from the index. However, if you believe the display of this information violates your rights as a copyright holder, you can file a formal legal complaint (Google 2012b).

An example of this metadata-only view is shown in Figure 50.



Figure 50: *Harry Potter and the Sorcerer's Stone* (1998), No Preview available on Google Books (screenshot by author, July 2013)

These views represent perhaps the most visible and direct instance of social factors influencing technology in any of the four cases in this dissertation: the fraction of the text that Google shows is inextricably linked to the legal context in which the project is situated – a socially-defined environment. In their diversity and criteria for application, these book views depict the complexity (and frequent irrationality) of U.S. and international copyright law in a much clearer and more publicly accessible way than any law review article could accomplish. They do not say, but *show* the extent to which digital access to information is complicated and often hamstrung by laws that extend control over cultural materials well beyond their most useful and profitable timeframe. Snippet view in particular – that “*bad user experience*” (P8) – starkly reveals the extent to which we do not know who owns what rights within the literature of the twentieth century (or indeed, whether those rights are even owned at all). Copyright law is generally quite poorly understood by the average individual, despite its pervasiveness in so many of our daily activities; the different ways of looking at books through Google, while far from a remedy to this lack of understanding, at

least provide an entry point for users to begin to ask the relevant questions: why can I get to this book, but not that one? Why are there only snippets for one edition of *Alice in Wonderland*, but full text downloads for another? The extent to which the creative output of the twentieth century remains unavailable to build upon – and indeed, often unavailable even to access (Heald 2013) – tends to remain invisible to the public mind as long as the materials themselves remain cloistered in library stacks. Instantiating those restrictions within the Google Books interface – translating and inscribing legal restrictions into public access structures – allows the users of that service to directly experience just how much has long been hidden, and remains partially hidden, within the written cultural record of the world.

4.2.2 Library Hosting and Serving

Of course, Google is not the only entity with the right to host and serve these files; each library partner has the right to provide access to scans made from their collections as well.¹⁸⁸ Without going into as much detail as in the above description of the Google Books interface, it will be useful to note some of the other avenues through which users can find and access books scanned by Google, and some of the different strategies libraries have employed for hosting those files on the backend.

The first books went live via Google in 2005 (Last 2007); access via the libraries came somewhat later. Two of the earliest examples of libraries providing access to Google-scanned books came in 2006 and 2007, at the University of Michigan and NYPL, respectively (Albanese 2006a, Quint 2006b, Crawford 2007, New York Public Library n.d.). Both of these institutions have altered the way they provide this access since that time (and thus, no screenshots are possible), but a brief description of each will nonetheless be useful here, to illustrate some of the different approaches different libraries chose to take.

The University of Michigan's service, MBooks, was launched in September of 2006 (Quint 2006b, Young 2006b), and in retrospect, seems clearly to have been a direct precursor to HathiTrust, which offered very similar set of features when first introduced.¹⁸⁹ Specifically, MBooks provided an alternate, independent interface for searching books scanned by Google from UM, and for reading the full text of those in the public domain, with a few more bells and whistles than Google was (or is) offering. As *Library Journal* reported at the time of the launch,

Under the program, the online catalog will point to an "MBooks" button that, using Google's scans, allows patrons – in the United States for now – the option

¹⁸⁸ And of course the Partner Program participants can do whatever they like with the content they own; to my knowledge none of the publisher partnerships involved any sort of exclusivity – but then, those contracts are generally not public, so it's difficult to say.

¹⁸⁹ That functionality has extended quite a bit in the five years since the launch of HathiTrust (Wilkin 2008).

of keyword searching within a volume and retrieving the number of times a search term appears per page. Users can then virtually “flip through” the selected work using a page-turning function. It also has enhanced features such as zooming in or out on pages and rotating images (Albanese 2006a).

A few points are notable about MBooks. First, it was the earliest (successful) attempt to integrate the Google book scans into a library OPAC – a process which, given the scale, turned out to be a fair bit more complicated than the institution’s previous experiences with digitization had led them to expect (Powell 2008). And second, UM was the first library to independently host the Google scans: when a user clicked on a link to an MBook, that link pulled the content from University of Michigan servers, and not from Google. Such hosting requires significant technological capacity and ongoing investment, and accordingly remains quite rare among Google’s library partners.

The New York Public Library’s implementation of access to Google Books through its catalog illustrates the more common way of integrating those collections into library services. That is, rather than constructing an entire independent interface like MBooks, the NYPL initially set up links from their catalog directly to the Google versions of the scanned books, at books.google.com (Anonymous Syracuse Student 2009, New York Public Library n.d.). And although the NYPL’s information page for its Google partnership still expresses an ongoing “intention to eventually provide direct access to copies of the digital files [they] receive from Google” (New York Public Library n.d.), this now seems less likely than it might have a few years ago. At this point, as it turns out, the scans the NYPL catalog links to are no longer hosted by Google – they are hosted by HathiTrust, in which NYPL is a contributing partner (HathiTrust Digital Library n.d.-b). Given the obvious efficiencies to be gained by pooling this hosting responsibility at a central, yet independently and communally governed, point, there would likely be little to gain for institutions like NYPL in creating their own independent services.

Moreover, it is worth adding that HathiTrust (Figure 51) was also the entity that replaced MBooks – and quite thoroughly so. All of the functions that MBooks used to perform for the University of Michigan now flow through HathiTrust,¹⁹⁰ and its functionality is also available to other participating institutions and their users. This arrangement has three major advantages for the libraries on its face: first, as noted above with regard to NYPL, it absolves each institution of sole responsibility for managing hosting and interface maintenance for their digital book collections – whether scanned by Google or by others. Second, it allows these institutions to pool their digital book collections to make a much larger corpus of works than would be available to any one of them acting alone. And finally (and not at all inconsequentially), it provides a centralized repository for these collections

¹⁹⁰ Indeed, examining the project documentation on the UM library website (<http://www.lib.umich.edu/michigan-digitization-project>), it almost seems as though MBooks never existed at all.

that is *not* housed within a for-profit corporation, whose attitude toward providing digital book services could change at any time. More than perhaps any other entity, HathiTrust puts the lie to those critics who continue to claim that Google is locking up culture into its own exclusive walled garden. Although HathiTrust is not without its limitations, it goes a substantial distance toward ensuring the perpetuation of the broad-scale corpus as a whole as time goes on, regardless of what happens to Google.

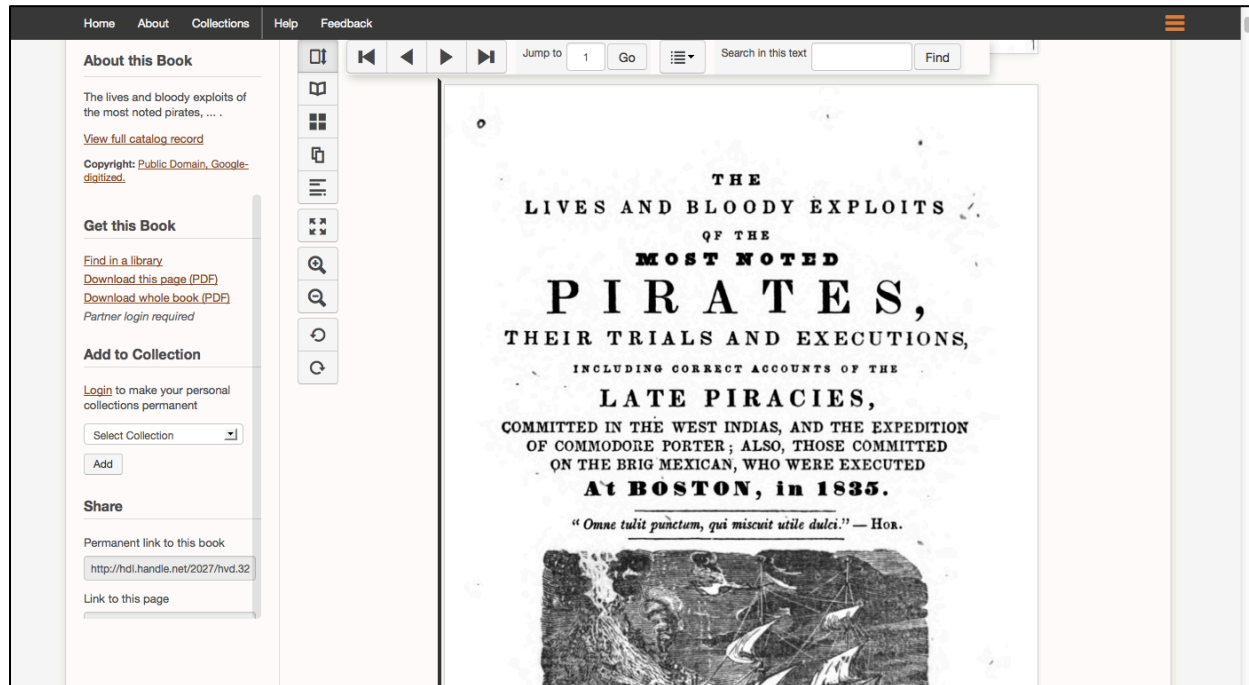


Figure 51: *The Lives and Bloody Exploits of the Most Noted Pirates* (1836) on HathiTrust (screenshot by author, July 2013)

All of the hosting and serving solutions developed by libraries, like Google's set of book view interfaces before them, have elements that can be traced to the sociotechnical milieu of their creation. The more technical piece has already been noted: many libraries simply lack the server and staff capacity to handle their own hosting for these materials. But the social piece bears mentioning as well. That is, although the libraries had a considerable amount of freedom regarding what they did with their copies of the GBLP scans, there were also a few significant limitations. First, like Google's interfaces, the libraries' had to embody an interpretation of copyright law. Within HathiTrust, for example, works that have been determined to be in the public domain are readable online in full, while works that are – or may be – in copyright are only available in "Limited View," which shows no part of any page – only a list of the pages where one's search term appears, and its frequency on each of those pages (Figure 52). The differences between Hathi's "Limited View" and Google's "Snippet View," moreover, highlight the second major socially-imposed limitation on the library interfaces. That is, the libraries are not allowed to build services that directly

compete with Google’s services. As such, whatever access the libraries provide to in-copyright works, it must not only comport with copyright law; it also must not compete with Google’s snippet view. And in the case of public domain works, HathiTrust does not allow PDF downloads with the same freedom Google does – one has to be logged in at a partner institution in order to download the full text of a book. Doing otherwise would likely violate the libraries’ non-compete agreements, and could potentially violate their agreement to make bulk downloading difficult as well, depending how it was implemented. As these examples illustrate, an entire network of social relationships and power dynamics have been infused into the digital book interfaces offered on both sides of the GBLP.¹⁹¹

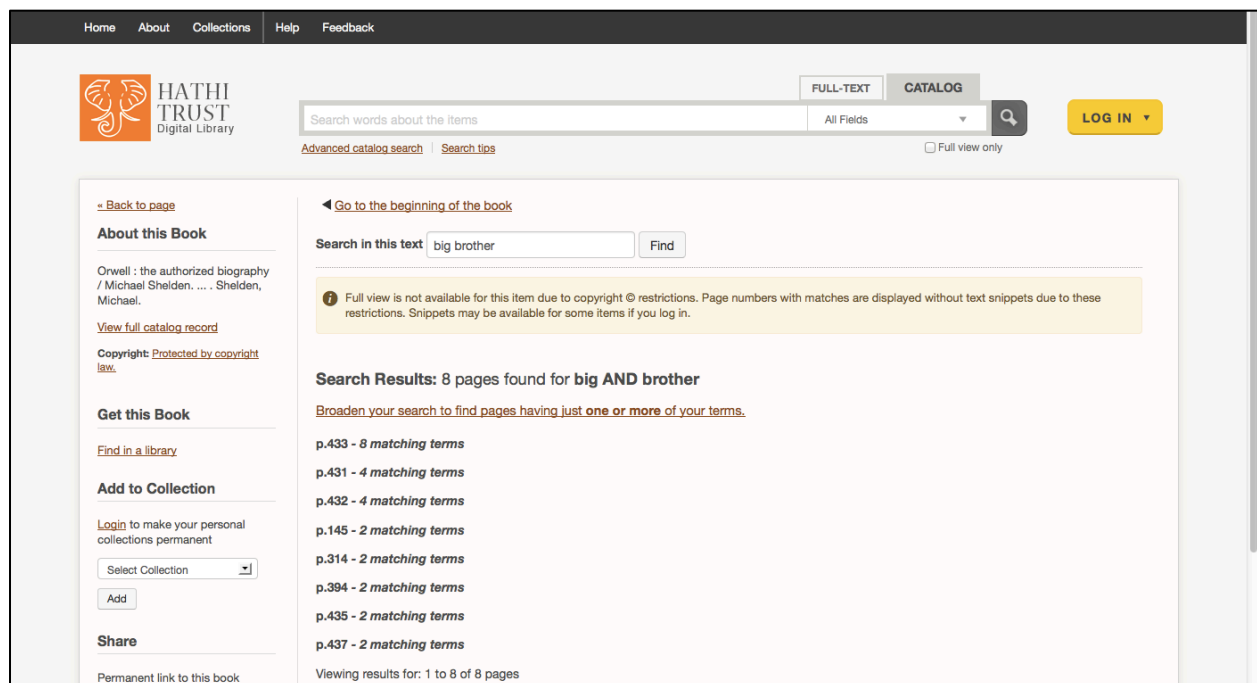


Figure 52: A search for “big brother” in *Orwell: the Authorized Biography* (1991) in Limited View on HathiTrust (screenshot by author, July 2013)

4.2.3 Standards

The issues regarding standards in Google Books – or, for that matter, in any of a number of large scale digitization initiatives – could fill another entire chapter, if not another entire dissertation. Particularly with regard to image quality and metadata, this topic is likely second only to the legal issues in the amount of vocal criticism and frustration it has inspired (e.g., Baksik 2006, 409-10, Duguid 2007, Townsend 2007, Jackson 2008, Nunberg 2009a, Nunberg 2009b). However, in the context of this chapter, the discussion of the

¹⁹¹ And this without even getting into the heavy references to Western library and bookstore metaphors within both the Google Books and HathiTrust interfaces – a design move that Duncker’s research has shown to be less than fully translatable across intercultural boundaries (2002).

standards issues in Google Books will be limited to a brief explication of way in which the service's demonstrated problems with both image quality and metadata are linked to two core elements of the project and of Google itself – that is, beta mentality and massive scale.

4.2.3.1 *Image Quality*

There is no denying that some image quality issues exist in Google Books. Some scanner-fingers are visible. Some text is cut off. Some pages are missing. These things are all true. However, as time goes by, they are becoming less true. One of the key advantages of Google's scanning technology, described above, is that it gathers a great deal more image data than is strictly necessary to present a page view onscreen – and as such, even after the scanning is finished, and the physical books have been sent back to their respective library shelves, the digital images can continue to be sharpened and improved through ongoing image processing interventions. As one library partner explains,

what they capture is a raw, unprocessed image and what most systems do is process an image. So, they were able to go back and iterate over previously captured images to get higher quality later on. So people would say "The scanning is bad." And what they were concluding was the processing was bad. The scanning was fine. And those same images today look good because the processing went back and got something good out of them. But it's a very different kind of approach or mentality to scanning than what typically happens. You do a lot of stuff to make sure things are right and you end up with slower throughput. They got faster throughput and they end up with a high quality product because they were able to invest in processes that improved it over time (P8).

Indeed, since the marked slowdown in scanning in 2011, this kind of post-processing has become an ever-greater focus for Google's Books-Project engineers (P10, follow-up conversation, July 2013). And in my own use of the books, I have recently noticed a difference, specifically in the deletion of some of the scanner-fingers (though the shadows surrounding them often remain).

The initial quality standards for Google's scanning – at least for the files they would deliver to the libraries – seem to have been developed primarily through negotiation with the three influential librarians mentioned earlier: John Wilkin at Michigan, Catherine Tierney at Stanford, and Dale Flecker at Harvard (P8, P17). Working collaboratively alongside the Google engineers involved, these individuals and their respective teams worked to develop benchmarks for acceptable quality in the images the libraries would receive. Notably, the focus in these negotiations was quite specifically on end product, and not on micromanaging the scanning process; as one librarian explains, "*What we needed was preservation quality images to do the work that we do and we could give them specifications on that and say 'How you get there is your business and your specialty'"* (P8). Additionally, even after these benchmarks were in place, many (if not most) library partners also implemented quality review procedures, in which a sample of incoming scans were continually examined

to ensure that Google's output continued to meet or exceed expectations (e.g., P8, P17, University of California and Google 2006, University of Wisconsin and Google 2006). This dual approach, of setting benchmarks and then spot-checking the end product – as opposed to striving for ideal scan quality on each and every individual page – allowed for the scanning to proceed at much greater scale and speed than would have been attainable otherwise.

The Googlers I spoke with, moreover, placed a strong emphasis on the iterative nature of the project – the extent to which the company was willing to try something, release it to the public, and then go back and improve it if and when problems were found. They also emphasized the close and necessary relationship between the project's massive scale and this process of iteration. One described the company's approach to book scanning through an analogy to digital image formats – traditional JPEG vs. progressive JPEG – explaining that,

The way images used to render on the screen, it would be like a raster. And it would start from the bottom, blop blop blop blop blah [making a line of pixels in gestures]. And when it rendered a pixel, it got the pixel perfect. And all its neighboring pixels perfect. But it took a long time till the full screen was done. Whereas progressive JPEG, it gets it blurry, and then sharper and sharper and sharper and sharper. So the way I describe it is the way scanning had been doing before was kind of like traditional JPEG. If a pixel was...the uh, page, being scanned at high quality, let's call that a pixel, ok? And now what you do is, you'd have this, and, and so, every page, in every library, in the US, is a pixel. So you have billions and billions and billions of pixels. And you'd start filling this one in and this one in and this one in and this one in [more pixel-making gestures]. Yet, it was gonna take...oh, a hundred and twenty years, until you could see the whole image. Now what Google Books did was say, hey, this is the wrong way to think about it, and it's actually economically inefficient. Because the cost of doing a scan relates to your confidence of that scan being perfect.

So in fact if you want a 99% confidence it'll cost X. If you want a 99.9% confidence, it'll cost sometimes X squared. OK? If you want a 99.999% confidence, it, it's 8 X squared, it just keeps going. In effect each extra "9" often two, three, four, five X the cost. It at least doubles...with each, with each extra "9". OK? So if- in this strategy, how much would it cost to get everything perfect? Well, it's gonna cost you a lot. OK. In the Google Books strategy approach, it was saying, look, we're gonna go in and scan at scale. We have a process to detect errors. It has many cycles (P4).

By taking this approach of shooting for very good (rather than perfect) quality with some tolerance for errors, Google was able to both greatly reduce the costs of its scanning and increase its speed, leaving the problem of perfecting the images to a later stage in the project. And, as this same participant continued, this strategy of tolerating some level of error likely allowed Google to produce a vastly greater number of "perfect" scans than would otherwise have been possible:

it's a different strategy. Right? And it's a strategy that...yes it, it, there are pages that

are bad? Right? And so...great. Suppose instead of scanning 20 million books we had scanned 10 books – 10 mil- sorry, 2 million books, and hadn't scanned the other 18 million books. OK? Let's now assume – and this isn't a stat that I know – but let's assume right now that 80% of our books are perfect. OK? So we scanned 20 million books, and 16 million of 'em are perfect. And 4 million of 'em aren't. So would we had rather scanned 2 million books and gotten 'em all perfect, or 16 million books, and only gotten 16 million of our 20 million books perfect. Well I woulda rather scanned 16 million of our books, or even for that matter, let's assume only fifty percent of our books are perfect. I'd rather get 10 million perfect books and 10 million that all need to be re-scanned than to take an approach where I only got 2 million to begin with (P4).

So, in sum: there are certainly some image quality concerns with regard to Google Books. Not all of the scans are as high-quality as they might have been had they been done by libraries, or even by the Internet Archive. However, the vast majority that I have encountered in my own exploration of the site have been perfectly usable for reading and research, and at least for the time being, Google seems committed to continuing to improve the quality of those images through post-processing (and even if they do not, there is nothing to stop the libraries from improving their copies, or re-scanning some volumes if necessary). This seems to me a clear case where the perfect ought not get in the way of the good – after all, *good* scales; *perfect*, not so much.

4.2.3.2 *Metadata*

The metadata problems on Google Books are arguably of greater concern – and they are legion. In 2009, a spirited and detailed conversation about these issues took place on Geoffrey Nunberg's *Language Log* blog (2009a), and despite its age, both Nunberg's analysis and the commentary that followed it remain quite relevant today. In the initial post, Nunberg avers that Google's metadata "are a train wreck: a mish-mash wrapped in a muddle wrapped in a mess," and specifically identifies multiple errors that he has found in that metadata – especially in the publication dates, but also in the subject headings¹⁹² and elsewhere. Ultimately, he concludes that these errors "are endemic" in Google's collection and, while acknowledging the company's "sincere desire to get this stuff right," goes on to assert that the strategy they have used for the imaging issues, of fixing errors on a case-by-case basis "isn't adequate [for the metadata]: there are simply too many errors" (2009a). My

¹⁹² Nunberg also criticizes Google's decision to utilize the bookstore-oriented BISAC subject codes to classify its collection, rather than a more extensible, library-oriented system like the Library of Congress subject headings. As he points out:

The BISAC scheme is well suited to organizing the shelves of a modern 35,000 foot chain bookstore or a small public library where ordinary consumers or patrons are browsing for books on the shelves. But it's not particularly helpful if you're flying blind in a library with several million titles, including scholarly works, foreign works, and vast quantities of books from earlier periods. ... In short, Google has taken the great research collections of the English-speaking world and returned them in the form of a suburban mall bookstore (Nunberg 2009a).

own experiences using the service, though admittedly unsystematic with regard to metadata evaluation, generally lead me to concur with Nunberg's assessment: many dates are very wrong (e.g. dated before the birth of the author), the metadata on serials are all but useless, and the subject headings are usually unhelpful, and often utterly perplexing or even comical.

Interestingly, in a rare example of transparency and engagement with critics on Google's part, Jon Orwant, then the Engineering Manager for Google Books, posted two substantive comments in response to Nunberg's critique shortly after it first appeared. In these comments, he largely accepts Nunberg's criticisms, and provides some valuable – if partial – insight into the reasons for them. To begin with, he writes:

Geoff refers to us having hundreds of thousands of errors. I wish it were so. We have millions. We have collected over a trillion individual metadata fields; when we use our computing grid to shake them around and decide which books exist in the world, we make billions of decisions and commit millions of mistakes. Some of them are eminently avoidable; others persist because we are at the mercy of the data available to us (Nunberg 2009a, comment by Orwant, September 1, 2009).

He then goes on to explain that Google has patched together its book metadata from “over a hundred” different metadata sources, and that many, if not most, of the errors that appear on Google Books come straight from those diverse sources. For example, responding to Nunberg's observation that a large number of books had been erroneously dated 1899, Orwant reveals that a Brazilian metadata vendor had been using “1899” as a placeholder where it did not have any date information at all – an error made outside of Google but ingested directly into Google's system. Orwant also describes the origins – both outside of Google and within it – of several other dating errors identified in the original post:

The 1905 date for the Drucker book was courtesy of a New Jersey metadata provider, which used 1905 in the same way that the Brazilian provider used 1899 (this, by the way, is a large part of the reason why there are so many books purportedly mentioning "Internet" prior to 1950). The 1900 Virginia Woolf date came from a British union catalog that has multiple MARC 260.c fields, some with the correct date and some without, but the 1900 field also occurs in the record's MARC 008 field. A time-traveling Tom Wolfe wrote *The Bonfire of the Vanities* in 1888 rather than 1988 because one of our all-too-human humans miskeyed the date. Henry James wrote *What Maisie Knew* in 1848 rather than 1897 because a French union catalog tells us so. Four bad dates, four different causes (Nunberg 2009a, comment by Orwant, September 1, 2009).

The problems in subject classification, Orwant goes on to claim, are attributable to a similar set of causes; he emphasizes that Google does not pull dates or any other metadata from

OCR (although in further comments, Nunberg questions this claim), and that if there are errors, they were made by humans, either at Google or at their metadata providers.¹⁹³

These examples, though incomplete in their coverage of the problems with Google's metadata, should be sufficient to support a stipulation at this point: the metadata on Google Books is not good. Hopefully, like the image quality, the metadata will improve over time; but for now, it is unacceptable. But rather than going into further detail on the specifics of the problems, it will be useful here to delve a bit more deeply into the social and technical contexts that bred many of these problems in the first place. That is, as explained by Orwant, many of the errors in the Google Books metadata emerged directly from Google's approach to gathering that metadata, piecing together over a hundred sources of apparently widely variable quality. But why did Google feel the need to do that? Could they not have simply gathered up the already perfectly adequate catalog records from their library partners, and used those? Sadly, things are not so simple. As it happens, Google *does* gather the records from library partners – but they are not allowed to display very much of the data in those records, because doing so would violate the *libraries'* preexisting agreements with another corporation – namely, OCLC (e.g., P10, Coyle 2008).

OCLC is a nonprofit corporation (at least nominally) which, among other things, is responsible for running WorldCat, a union catalog through which a vast percentage of library cataloging work flows. Within libraries, many records are downloaded from WorldCat rather than cataloged independently, because doing so both increases efficiency and keeps the library's catalog records in sync with records at other libraries, which facilitates functions like interlibrary loan. Participation in WorldCat is by paid subscription, and that subscription comes bundled with restrictions regarding what libraries can do with the records they have downloaded from WorldCat. Orwant's comments hint only vaguely at Google's difficulties in wrangling with OCLC over records; librarian Karen Coyle, however, states the issue more directly. On her own blog, she asserts:

I happen to know that Google receives a full bibliographic record with each book that it digitizes. I have no idea what they do with that data because it doesn't appear on the screen in Google Book Search. What is significant is that the quality data that has been created by libraries is still essentially invisible to most users of the Internet (2008).

And further, she goes on to explain, this higher-quality data likely remains invisible at least in part because of limitations placed by OCLC on the specific metadata fields that libraries are allowed to share with other entities for display purposes. By way of illustration, Coyle

¹⁹³ As for the decision to use BISAC, Orwant says only “we thought our end users would find it useful” – a claim to which Nunberg immediately responds, “The question is, why did you think end-users would find this useful? Which end-users did you talk to about this? I don't think you'd find a whole a lot of scholars who would embrace the idea of using the BISAC classifications in place of other library classification schemes.”

goes on to describe a then-recent example of records “amputation,” in this case taken from a University of Michigan data release, comparing the sparse set of metadata the university has been allowed to share to the much richer set that appears in the UM library catalog. Without delving into the nitty-gritty details of the MARC record fields involved,¹⁹⁴ it is worth noting that among the omissions in the truncated records are pieces of the authorship data, the place of publication, and all but the first segment of the library subject heading (Coyle 2008).¹⁹⁵ Assuming that this sort of truncated metadata is all that Google is allowed to use from most of the library records it receives,¹⁹⁶ their quest for supplementary data sources becomes more comprehensible – as does their rate of errors and omissions. Libraries signed away the right to fully share their metadata long before they signed on to work with Google; Google’s scanning simply made the unintended consequences of those contractual restrictions more visible (by several orders of magnitude).

As a final note on Google’s metadata, one last question should be addressed: namely, given that Google is offering full text search, why do we even still care about the metadata? Can’t we just go forth and find things in books the same way we have grown accustomed to finding things on the web? And the answer here is yes, in part. To the extent that a reader is looking for a particular phrase in a book, or books that use particular words in the text, Google’s model of search and retrieval should work quite well – much better, in fact, than any pure catalog system ever could. However, there are two key issues with this approach for books. One, to be discussed below, is the illogic of applying Google’s existing Web retrieval and ranking algorithms directly to books – is the book most frequently mentioned on web pages (or perhaps in other books?) truly going to be the most relevant result for a scholar researching some obscure topic? And the second is that high-quality metadata allows for forms of retrieval that are simply not possible through pure text mining. Subject headings are a classic example: a book of poetry might contain nothing but works about love, yet never actually mention the word “love;” the “aboutness” of an item cannot always – or even often – be ascertained mechanically or algorithmically. (Humans are still useful for some things.) Or take dates. A book on the Civil War would likely contain numerous references to years in the 1860s, regardless of when the book itself was written. If you wish to execute a linguistic analysis of how depictions of the Civil War have changed over time,

¹⁹⁴ For those who would like to delve into such issues, they are well explained in Coyle’s original post, at <http://kcoyle.blogspot.com/2008/05/amputation.html>

¹⁹⁵ Regarding the last of these, moreover, Coyle notes:

The subject headings have been rendered entirely useless. As we know, the 6XX \$a is not the top of some logical hierarchy, but is idiosyncratically the first term based on some rather complex rules. So in the first record we lose "United States" because it is the second term, but in the second record we get only "United States" and lose all references to "Marine Corps." which is the actual topic of the item (2008).

¹⁹⁶ The records that the libraries still create in-house would presumably still be theirs to share; just not the ones downloaded from WorldCat. Teasing out which are which may be another question, however.

you will need an accurate publication year for each book, and not just a harvesting of the years discussed between its covers.¹⁹⁷ Full text search or not, the metadata is well worth getting right.

4.2.4 Invisible structures

Finally, there is one further class of structures whose specific shape is not, and likely never will be, fully knowable by anyone outside of Google: that is, the mechanisms through which books are ranked and presented for browsing within the Google Books interface. It is common knowledge that Google's search algorithms are among its most closely held trade secrets. And in the web world, there are excellent reasons for this: in particular, the fact that entire industries have grown up around trying to game that algorithm – to spam Google searchers and “bomb” the results with irrelevant (though occasionally amusing) links (e.g., Granka 2010, 366). And beyond this, of course, keeping the search algorithm secret also keeps Google's competitors from capitalizing on the company's innovations in that area.

Still, we actually do have *some* sense of how Google ranks web pages. The fundamental idea behind its PageRank algorithm – utilizing a mix of popularity and authority as a basis for search rankings – was published by the founders in 1999 (Page, et al.). Combined with various other signals, including linguistic markers and user behavior (e.g. clicking links, reformulating queries, even cursor movements), PageRank still forms the core of Google's web search today (Granka 2010, 366-68). As various commentators have pointed out, however (e.g., Madrigal 2010, Vaidhyanathan 2011, 171), this algorithmic system is inherently “webby” – it doesn't obviously translate to searching books and other born-analog materials, which lack the hyperlinks and other structural features upon which web search relies.

So how does Google rank books? Some indication is given by a patent filed by the company in 2010 (Petrou, et al.), which lays out a number of signals within books¹⁹⁸ that Google has deemed useful for determining relevance within a search context. In particular, the patent

¹⁹⁷ Indeed, this is a key question that merits answering with regard to Google's “n-gram” viewer (<http://books.google.com/ngrams>) and the scholarly publications that have emerged from it (e.g., Michel, et al. 2011). That is, the n-gram viewer dataset was compiled from about 5.1 million books scanned by Google, mainly from the library project (Michel, et al. 2011). Its functionality – showing how words or groups of words have shifted in usage over time – relies fundamentally on accurate metadata, and especially on accurate dates and OCR. And its creators assert that books were included “on the basis of the quality of their OCR and metadata” (Michel, et al. 2011). Yet, the creators themselves assert that “The corpus cannot be read by a human. If you tried to read only English-language entries from the year 2000 alone, at the reasonable pace of 200 words/min, without interruptions for food or sleep, it would take 80 years” (Michel, et al. 2011). So, given that vastness, how was the quality of the metadata and OCR assured? How did that quality assurance manage to scale to 5.1 million books? And if it did scale that well, why can it not be scaled to 20 million?

¹⁹⁸ Both conventional books like novels and text books and also what the patent calls “non-traditional books” – e.g., “newspapers, magazines, journals, pamphlets, web pages and other electronic documents, and the like” (Petrou, et al. 2010, 2).

describes how Google tracks certain “entities” – things like the names of places or people, dates, and events – and uses them to aid in its rankings. As concisely explained on an SEO¹⁹⁹ blog, *SEO by the Sea*, Google derives four major signals from such entity-tracking:

1. How much information about a specific entity is included in the book and where. ...
2. Whether there are third party references point to a particular book and to its mention of specific entities. ...
3. Whether or not the sections of a book that include that entity are accessed more than other sections of the book. ...
4. How frequently that entity is mentioned in the book compared to how frequently the entity is mentioned in other books in the collection of books (Slawski 2011).

So, for example, a book with the words “Robert E. Lee” on its title page might rank higher in a search for the Civil War General than a book that mentions Lee’s name only in the text – although if a book’s text mentions Lee far more frequently than most other books, his name not being on the title page might not make as much difference (#1 and #4, above). And should the book’s reviews (or articles citing the book) also mention Robert E. Lee, that would tend to signal a higher rank for those terms as well (#2 above). And if Google Books users tend to gravitate most toward the chapter of the book that contains the most mentions of Lee, that would provide a further indication of its utility for that topic (#3 above).

Additionally, Madrigal reports based on discussions with Google engineers that these document characteristics are not the only considerations: Google’s algorithms also “take into account web search frequency, recent book sales, the number of libraries that hold the title, and how often an older book has been reprinted” (2010). These are all essentially measures of a book’s popularity, though also to some extent its importance (library holdings) and ongoing relevance over time (reprintings).

For the most part, this seems like a perfectly reasonable set of criteria to use for ranking books in a largely consumer-oriented setting. Still, these are clearly not *all* of the criteria Google uses. And even cursory use of the books.google.com search engine reveals much to be desired in the site’s rankings. For example, a search in Google Books for Phileas Fogg – the main character from *Around the World in 80 Days* – turns up a list that ranks both the Wikipedia entry for the character and a 1993 spinoff of the book (No Preview available) above Verne’s actual novel (which, having been published in 1873, is available in full text). These results (as well as one with a more colorful title) are shown in Figure 53.

¹⁹⁹ SEO = Search Engine Optimization. Essentially the polite name for the aforementioned practice of trying to game the algorithms of search engines to get your product or company placed higher in the results.

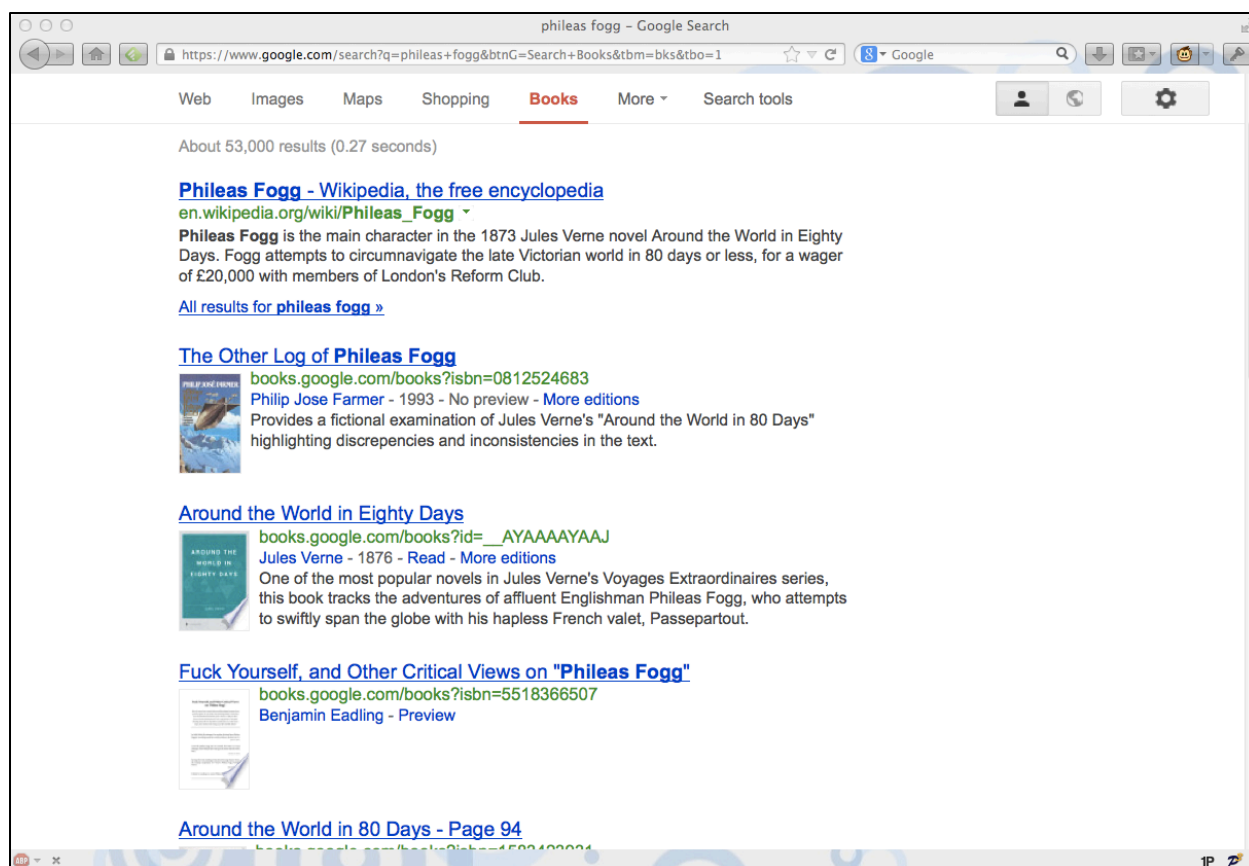


Figure 53: Top four results for a Google Books search on “Phileas Fogg” (screenshot by author, August 2013)

To be fair, it is perfectly possible that a searcher might appreciate the Wikipedia entry, if all they wanted was basic information about the character and not the actual book. However, this list – and its two public domain hits out of the first ten – begins to indicate a broader issue that I have encountered when searching Google Books – that is, something in the algorithm seems to privilege books that are not available in full view over books that are. For example, a search for “Snow White” – a traditional German folk tale popularized by the Brothers Grimm in the early nineteenth century (and thus a part of the public domain many times over) – brings up more than a full page of Limited Preview and No Preview editions before finally showing one that can be read online. And in fact, that full text result – a 1921 edition of *Grimm’s Fairy Tales* scanned from the New York Public Library – is the *only one* in the first *five pages* of hits; none of the other 49 top-ranked results can be read online for free. To a casual searcher unfamiliar with copyright, it could well appear that Snow White is not in the public domain at all. And while it is true that one can limit any given search to full text only, such tinkering should not be an absolute requirement in order to find public domain versions of public domain works, or even to discover that such versions exist at all.²⁰⁰ Non-transparency regarding particular technological structures can be an

²⁰⁰ After all, it is a rare searcher who goes beyond the first page, much less beyond the fourth.

understandable strategy for a technology company; however, when that secrecy leaves open the possibility of so many different kinds of biases in areas so fundamental to information access and the rights of the public, it remains well worth questioning.

5. Conclusion

The Google Books Library Project is a massive endeavor that has changed the world of books and libraries forever, for better or worse. Motivated by a complex mix of personal interests, institutional missions, and dovetailing of pragmatic goals on the part of libraries and Google, the project has facilitated the scanning of books at a scale never before imaginable – more than 20 million at last (public) count – and has opened up the ability to search within those books to a user base far beyond the reach of any physical library. Many significant questions clearly remain regarding the ways in which the procedures it has pioneered and the structures and resources it has created will ultimately affect the individuals, institutions, and social systems into which they have or will come into contact. In particular, concerns surrounding reader privacy, intellectual freedom, the implications of poor metadata for “big data” work in the digital humanities, and the potential impacts of biases in the underlying algorithms all form important areas for further inquiry and critique. However, these issues are a topic for a career; the purpose here has simply been to lay a foundation for thinking about them more critically, and with a stronger grounding in the facts of the project’s unfolding.



To conclude this chapter, I would like to offer a brief anecdote from my fieldwork.

In wrapping up one of my interviews at the University of Michigan Libraries, I asked one participant whether he thought that *“if Google Books died today, it’d still be a good thing for the University of Michigan.”* His response? Of course it would. And, he explained, the reason he thought so was:

We needed to change the conversation of what was happening in the library community, and this, this allowed us to change the conversation about what’s happening in the library community (P8).

As noted much earlier in this chapter, the conversation surrounding large-scale digitization in libraries used to be all about “whether” – whether it was worth the cost, whether anyone would use it, whether a sufficient number of institutions could be gathered to participate – but since the GBLP was announced, the dialogue has entirely shifted to considerations of “how,” “when,” “by whom,” and “how fast.” HathiTrust has millions of volumes in it, and is sharing its public domain holdings with the Digital Public Library of America – but before Google Books, neither of those projects were even a glimmer in anyone’s eye. I tend to agree with this participant – even if the project had achieved nothing else, this shift in the digitization zeitgeist would have been an outstanding achievement all on its own.

Whatever its problems – and I have raised several above – the GBLP has done a great service for the dialogue surrounding digitization in general. (And 20 million digital books make for a nice bonus.)

Case 4: Open Content Alliance (2005-2011)

The Open Content Alliance (OCA) has a dubious distinction among the cases chosen for this dissertation: that is, it is the only case that can reasonably – though not unequivocally – be said to have failed. In fact, during my very first interview regarding the OCA, the participant flatly told me “*there is no OCA*” (P1) – and remarkably, a later participant repeated this exact phrase: “*there is no OCA*” (P11).

Having been studying the OCA for years already by the time of these interviews, I was fairly shocked that this was the impression of the project imparted by some of its most influential leaders. After that first interview, I began to directly ask each participant for this case – three individuals affiliated with the Internet Archive, five affiliated with OCA-contributing libraries, and two affiliated with both IA and at least one contributing library – whether they believed the OCA existed. A few of them initially laughed – it does sound like an absurd question in the context of an interview *about the OCA*, after all – but then they would take a moment to ponder. Ultimately, the depictions of the project in these interviews began to feel like *Rashomon*, the Kurosawa film in which the same story is told from several perspectives, and the meaning becomes utterly different in each retelling. Taken together, as in that film, three alternate realities emerged:

1. “There is no OCA” – and if there ever was, it failed.

Among the interviews I conducted, three participants – the two already mentioned and one more – suggested that the OCA didn’t exist, and one of these even went so far as to claim that “*it never really did*” (P3). Despite their protestations of its nonexistence, however, all three of these participants also made reference to the OCA having *failed* – implying that at some point in time it *did* exist, but that it had ceased to do so by the time the interviews were conducted in 2011-2012. In the assertion of *failure*, though not of *nonexistence*, these three are joined by a fourth interviewee, who explains:

OCA as an entity is... essentially defunct. It's dead. I... obviously the benefit of the scanning will have a long life, I hope and trust, and believe, but I think organizationally it's not been as effective as, as early participants wanted it to be (P2).

Notably, all four of the participants in this narrative cluster were very deeply involved in the leadership of the OCA in its early years, though their affiliations to it varied. And all were at some point major proponents of the project, in very public ways. As such, the fact that this particular group of people have declared the OCA nonexistent and/or a failure seems quite revealing.

2. It’s a success – but “it” isn’t necessarily the OCA.

At the opposite end of the spectrum, two participants (P5, P7) averred that the project was absolutely succeeding. For example, one claimed:

I think it's doing a bloody good job, actually, I'm pretty amazed at how far we've come with such a tiny weeny little team. Um... and... ah... it's a big job, so you compare the Internet Archive effort with something like Google Books and you have, you know, um, a boutique butcher making his own sausages versus, you know, just this enormous kind of factory-farm situation. And, you know, they're sort of opposites in some ways (P7).

However, as the comment above hints, in both of these interviews there was a strong elision between the Open Content Alliance and the Internet Archive's book scanning efforts more broadly – a fact for which I take some responsibility as the interviewer. Both of these individuals were relative latecomers to the project, coming on at the IA end after much of the explicit "Open Content Alliance" discussion had already concluded. As such, their involvement with the OCA *per se* was peripheral at best; by the time they arrived, the focus had already shifted to related areas like the IA's physical book archive and Open Library. Within these two interviews, it was thus difficult to discuss the participants' involvement with the OCA, because in a sense, they had none; although they were influential in the ways that some pieces of the project moved forward, they missed the stage at which those pieces had cohered under the OCA's umbrella. Because of this, I focused these interviews on the pieces these participants were actually involved with, rather than dealing more explicitly with the OCA. As a result, although these two claim success, they arguably do so not for the OCA, but for the Internet Archive's ongoing book-related initiatives – some of which grew out of the OCA, but none of which are coterminous with that project.

3. It was a set of meetings and principles, and not an actual organization.

Finally, somewhere in the middle, three other participants – including the project's central figure, Brewster Kahle – describe the OCA as having been the articulation of a set of principles, rather than an actual organization. As Kahle puts it, *"It was a set of meetings and a set of principles. Um. To get things...get going"* (Kahle 2011). Building upon this idea, one library partner suggests that by the time of our conversation in 2011, the OCA had become *"an expression of people who agree with those principles as kind of, um, supported by Brewster. I don't think it... it's not separate from him"* (P14). And taking this view to a more philosophical level, another librarian muses:

I think it has... you don't hear so much about "Open Content Alliance" because it's part of the fabric. You know it's part of the library fabric right now. We're all... all the principles of the Open Content Alliance, people all nod in agreement and say "yep, yep that's the right thing to do." So I think it has become more of the fabric of what libraries are (P15).

Like the set of participants who declared the project a failure, these three were all very deeply involved in it from a very early stage. However, as this chapter will show – and as the two librarians just cited rightly note elsewhere in their interviews – the OCA was not *initially supposed* to be limited to a set of meetings and principles, though it did ultimately evolve in that direction. Rather, the OCA was supposed to take on the then-recently-

announced Google Books Library Project; to digitize books in the open, with support from Google's tech-world rivals; to prove that mass digitization could be accomplished collaboratively, transparently, and on democratic terms.

The historical synopsis below will lay out the details of the OCA's trajectory, to the extent that it can be ascertained from available evidence. But before delving into that story, here is my own brief take on the ontological question of the OCA: it did exist, but it does not anymore. Between late 2005 and early 2009, it facilitated a large amount of book scanning and at least an equal amount of open access (and anti-Google) advocacy. And then, around the middle of 2009, it quietly collapsed, apparently under the weight of a vacuum in both leadership and financial support. Beyond that point the Alliance split into several successor projects – digital lending and interlibrary loan, scan-on-demand services, a physical book archive, and more – both at the Internet Archive and at other participating institutions. In many of these cases, the transition was seamless; the activities evolved smoothly from OCA to not-OCA, with the term “Open Content Alliance” simply slipping out of use (at least for the most part) as the activities themselves continued.

Somewhat ironically, given the project's strong emphasis on transparency, the available data regarding the Open Content Alliance is considerably sparser than that regarding Google Books (or either of the historical cases). The OCA's blog cuts off without explanation in May 2010, and even before that, its posts had long ago ceased to discuss the Open Content Alliance as such (the last post to use the term appeared in July 2009). And even when that site was in regular operation, the level of detail it provided in describing the initiative and its various pieces remained relatively low. Beyond the OCA site, what information is available comes piecemeal from the various entities and individuals involved – principally at the Internet Archive, but also some of its library partners – as well as popular media. But these accounts are much fewer and further between than those about Google's project,²⁰¹ and equally as contradictory as the interviewee depictions of the project outlined above. The scattered state of the data on this case indicates two things for this chapter: first, there will be gaps in the story that no data was available (or at least findable) to fill; and second, the chapter itself will be all the more important for providing the first relatively comprehensive narrative of what actually might have happened to this once-inspiring challenger to Google in the world of book digitization. Figure 54, below, provides a concise timeline of the project; a more complete textual version, with citations to sources, appears as Appendix F. As in the previous chapter, different colors have been used to

²⁰¹ For scale, in conducting this research, I ultimately located more than 500 discrete sources on Google Books, but only about 200 on the OCA (counting both primary and secondary sources).

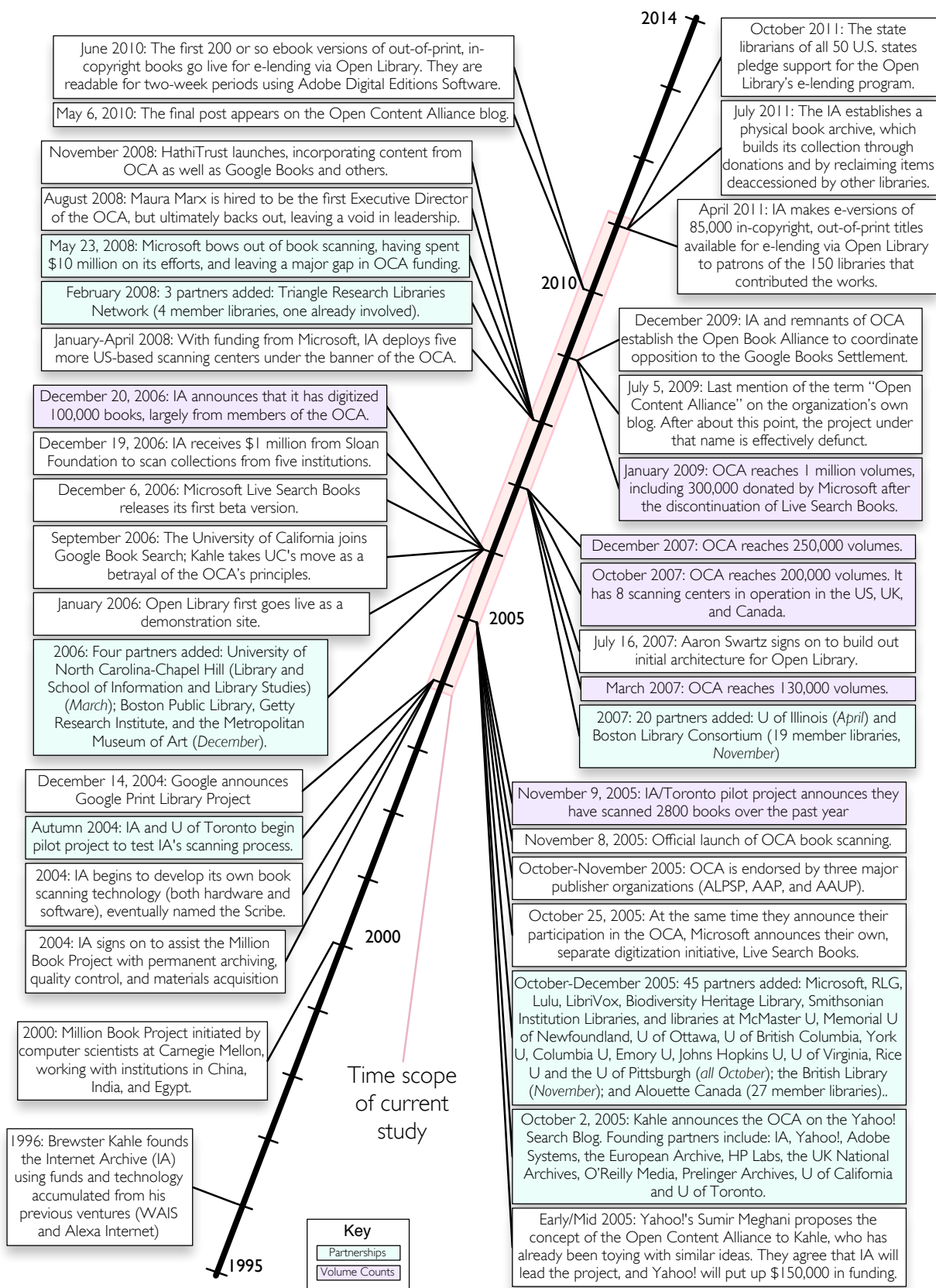


Figure 54: Timeline of the Open Content Alliance (citations in Appendix F)

highlight particular themes: additions (and subtractions) of partners in the project appear in blue; reported growth in the number of volumes scanned appears in purple.

1. Historical Synopsis

As indicated in the introduction, the OCA has not one history, but many. Still, it will be useful to lay out a roughly linear narrative of the significant milestones in its unfolding, to provide a basic backdrop for the analyses that follow. In what follows, I have done my best to note where disagreement exists over points of fact; however, for the sake of brevity and simplicity, this synopsis does not pick apart every contentious detail.

The Open Content Alliance was announced by Internet Archive Founder and “Digital Librarian” Brewster Kahle on the Yahoo! Search Blog on October 2, 2005 – about eleven months after the announcement of the Google Books Library Project. Within the press accounts at the time, it was typically depicted as Yahoo! and/or Microsoft’s attempt to compete with Google on book scanning (e.g., “Pulp Friction” 2005, “Yahoo Takes on Google” 2005, Auchard 2005, Chillingworth 2005, Hafner 2005, Hane 2005, Young 2005b, Goldsborough 2006, Young 2006a, Ashmore and Grogg 2008) – and several of these accounts credit one or the other of these companies with the project’s establishment, its leadership, or both. For example, an early *New York Times* headline about the project proclaimed, “In challenge to Google, Yahoo will scan books” (Hafner 2005), while the *Chronicle of Higher Education* announced, “Yahoo Works With Academic Libraries on a New Project to Digitize Books” (Carlson and Young 2005). But the media’s emphasis on the clash-of-the-tech-giants angle in these accounts was misplaced. As library and technology commentator Walt Crawford lamented at the time, “The competitive thrust may be necessary to make a newspaper story exciting, but I find it a bit unfortunate. OCA is and should be less and more than a ‘challenge to Google’” (2005). In fact, the OCA *was* both less and more than a challenge to Google – and perhaps more importantly, it was most significantly led and shaped neither by Microsoft nor by Yahoo!, but by the Internet Archive. And even more than that, it was led and shaped by the man behind its initial announcement – the Internet Archive’s Brewster Kahle.

In a very real sense, the OCA was Kahle’s creation. In my interviews, when asked the question “who played an especially strong role in shaping the OCA?” every OCA participant cited Kahle first, without hesitation. Two, in fact, cite him second and third as well – declaring (verbatim) “*Brewster, Brewster, Brewster!*” (P2, P14) – and two others credit him almost *exclusively*, only naming one or two other influential individuals when explicitly pressed (P3, P7). The press coverage aside, this is one of the most clearly established facts about the OCA: whatever else it was, it was Brewster Kahle’s baby. And given his centrality to the project, it makes sense to use his version of its origin story as a starting point for

describing its beginnings. In his interview with me,²⁰² he began the story in 1980, explaining that around that time, he

sort of fastened on the “why don’t we just build the library”... um... idea. And it required some things to happen first. First we needed sort of computers that could store enough, um... so I helped work on that. And then we needed a network so that you could actually get to the stuff, so I helped work on that. That was the late ‘90s. Then we needed software systems and a mechanism for publishers to make money, I worked on that in the early ‘90s. Um... by the time we sort of got the computers, the network, and the, and the publishers online, sort of ‘94, then I thought I could really start working on the library. Um... then... so... started with materials that were, um, in danger. So started with the... first with the web, then the next was television, um... and started working on books, um... in... 2001? Um... with the Million Books Project led by Raj Reddy’s project out of Carnegie Mellon with the governments of India and China (Kahle 2011).

This account of 25 years’ worth of incremental buildup toward an inevitable culmination in large-scale book digitization is basically supported by publicly available accounts of Kahle’s career trajectory. In the 1980s, Kahle worked on supercomputers at a company called Thinking Machines, contributing to making, as he put it in the passage above, “computers that could store enough” as well as “a network so that you could actually get to the stuff.” In 1992, he left to co-found WAIS, Inc., an early publishing and distributed search system. In 1995, he sold WAIS to AOL for \$15 million, and the next year, he simultaneously founded Alexa Internet – a web analytics company – and a nonprofit institution called the Internet Archive, which used Alexa’s archived web-crawls as the foundation for its archive of Internet content. Then, in 1999, he sold Alexa to Amazon for \$250 million, retaining the right for the Internet Archive to continue to receive its crawls. Shortly after that, he threw the Internet Archive’s support behind the Million Book Project,²⁰³ and that collaboration provided the initial inspiration for Kahle to develop his own scanning technologies, later named the Scribe system (Hogge 2005, "Frequently Asked Questions", Hardy 2009, "Brewster Kahle" 2013, "Technology", "Staff Bios"). Although the path from Kahle’s initial desire to scan books to the semi-realization of that desire in the Open Content Alliance seems unlikely to have been quite as linear or teleological as his presentation of it in the interview might imply, it does seem clear that Kahle can legitimately claim to have personally laid significant portions of the technological foundation upon which large-scale

²⁰² As previously noted, in his interview, Kahle explicitly “overstomped” the confidentiality statement I’d just read, thus allowing me to name him here.

²⁰³ The Million Book Project was an NSF-funded large-scale digitization project geared toward addressing a particular set of technological research questions, and not necessarily toward the provision of broad-scale access to information. It was initiated by a group of computer scientists at Carnegie Mellon, in collaboration with universities in China and India, the Biblioteca Alexandrina in Egypt, and other partners, in 2000 (St. Clair 2008, 152-53). At some point between its founding and the time when it faded away, in about 2008, it also became known as the Universal Library Project. Access to its scans are now provided through the Internet Archive Ebook and Texts Archive ("Frequently Asked Questions" 2007, Kahle 2011).

book digitization in general – and especially within the OCA – would ultimately come to rest.

Still, the account above only brings us to the beginning of the OCA timeline. The starting point of the project's more specific trajectory lies in the early autumn of 2004, when the Internet Archive began pilot-testing its book scanning processes at the University of Toronto. Notably, this IA/Toronto pilot test began at about the same time Google was pilot-testing its own processes at the University of Michigan, well before either the GBLP or the OCA was announced. However, the scale of the IA/Toronto pilot – like the scale of the OCA scanning that would build upon it – was much, much smaller than Google's project. From 2004 to 2005, that pilot project scanned 2800 books, at a cost of \$108,250 – about \$38 per book (Carlson and Young 2005, Crawford 2005, Bengtson 2006). No Google volume count is available for that precise period, but we do know that the GBLP reached 1 million volumes within its first *two* years (P4, P12), and can thus surmise that for the one-year period in question (2004-05), their count was already well into the hundreds of thousands – a few orders of magnitude greater than the IA pilot.

In early 2005, with the pilot scanning at Toronto already underway, and the Google Books Library Project announcement making waves in the digital information world, Kahle and others in and around the Internet Archive began to discuss the possibility of launching a broader-scale, open alternative to Google's project. Still, it took a proposal to the Internet Archive from Sumir Meghani of Yahoo! to coalesce and put a name to this project: the Open Content Alliance (P1). The Internet Archive readily agreed to lead the initiative, and Yahoo! contributed \$150,000 (earmarked for scanning in American Studies at the University of California) to get things started (P1, P3, Albanese 2005a, Johnson 2007, 5). Interestingly, however, despite Yahoo!'s pivotal role in launching the project – and despite the immense amount of credit the press gave them for leading the project throughout its lifespan – that \$150,000 seems to have been the only contribution Yahoo! ever actually made to the OCA: the book search system they promised to build at the outset of the project (e.g., Carlson and Young 2005) never materialized, and there is no definitive confirmation anywhere in the press coverage or interviews of their ever having provided any additional funding. Indeed, one librarian I interviewed sardonically noted that Yahoo! “*got more press for that \$150,000 than I think they got for any ever expenditure. It was the best marketing they ever did*” (P3).²⁰⁴

Ultimately, as previously noted, the OCA was officially announced by Kahle on the Yahoo! Search Blog on October 2, 2005. The founding partners in the initiative included the Internet Archive, Yahoo!, Adobe Systems, the European Archive, HP Labs, the UK National

²⁰⁴ No indication of *why* Yahoo! lost interest is given in either the interviews or any of the other sources of information for this case. The company simply seems to fade away as an active contributor, while leaving its name on the list of participants.

Archives, O'Reilly Media, the Prelinger Library and Archives, and the libraries at the University of California and the University of Toronto ("Consortium Forms OCA" 2005). Rick Prelinger, a personal friend of Kahle's who was already deeply involved in the digitization work of the Internet Archive through his film collection (the Prelinger Archives), was tapped to coordinate the project (P1, P3). Conceptually, the OCA was an immediate hit with libraries and publishers alike. Within two months, the project had been endorsed by three major publishing organizations (ALPSP, AAUP, and AAP), and had gained 45 additional partners: 41 libraries, one print-on-demand service (Lulu), one audiobook producer (LibriVox), one union catalog provider (RLG), and, of course, Microsoft, which would become the project's most significant individual source of funding ("RLG Joins Open Content Alliance" 2005, "MSN Search Announces MSN Book Search" 2005, Albanese 2005a, Auchard 2005, Carlson and Young 2005, Crawford 2005, Suber 2005). When Microsoft announced its participation in the OCA on October 25, they simultaneously announced the launch of their own, separate, non-OCA and non-Google initiative, Live Search Books ("MSN Search Announces MSN Book Search" 2005, Albanese 2005a, Crawford 2005). Within retrospective accounts of the project, there is often some elision between Live Search Books and the OCA – likely in part because after the conclusion of the former, it donated all its scans to the latter, and in part because they were in other ways interrelated – but during the time both were in operation, they were in fact separate projects (e.g., MSN 2005).

Over the next two years, the OCA's scanning proceeded apace – though still at a much slower rate than Google's parallel initiative. By the end of 2007, the Internet Archive and its partners had scanned 250,000 volumes, and had eight scanning centers in operation in the US, the UK, and Canada (Goth 2007, Hane 2007, Kahle 2007b, Ashmore and Grogg 2008).²⁰⁵ The IA began to develop the Open Library site, with open access advocate Aaron Swartz taking the lead in designing it as "a website with a page for every book, collecting everything we can find out about it from libraries, publishers, reviewers, and of course, book lovers" – including, where possible, full text of items scanned by the IA or other OCA partners (Kniffel 2008, quoting Swartz). Around the same time, Microsoft went live with its Live Search Books service, which was hailed as a useful alternative to Google – at least, while it lasted (it shuttered in 2008) (Guren 2006). And alongside these developments, more partners were added, the Scribe software was improved, and the Internet Archive was even officially declared a library in its own right (Albanese 2006c, "Boston Library Consortium and Open Content Alliance to Provide Digitized Books" 2007, Albanese 2007a, Hane 2007, Internet Archive 2007, Kahle 2007a, Kahle 2007b, A. McCoy 2007). One of the largest OCA partners added during this period, the Boston Public Library, was quite vocal about having

²⁰⁵ By a similar point in time, Google and its partners had reportedly scanned more than 5 million volumes – about twenty times as many as the OCA (Jacsó 2008).

rejected Google's scanning advances in favor of the OCA's more open alternative project (e.g., Albanese 2007a). And perhaps most promisingly, in December 2007 the IA received a \$1 million grant from the Sloan Foundation to support its scanning efforts on behalf of the OCA ("Sloan Foundation Grant Awarded" 2006, "Internet Archive Receives Grant" 2006, Internet Archive and Boston Public Library 2007).

Still, this time period was not all smooth sailing. In September of 2006, the University of California – at the time the OCA's largest single content contributor, and a founding partner in the project – announced that it was signing on with Google Books as well. From the UC's perspective, this move seemed only logical; as one librarian there explains, the sense was, the more externally-funded digitization UC could do, the better, regardless of the partner:

We didn't really care who [the first partner] was. It could've been Bill [Gates]. We also did Yahoo! I mean, Yahoo! was early in this game, uh, and I think they were our first, in fact Yahoo! was our first. And, uh, we did a deal – and that's actually Yahoo!, and the OCA was originally Yahoo!, UC, and Brewster, right. ... And then, um, and then Microsoft, we did Microsoft too. Microsoft was briefly involved, and I think we did a bunch of stuff with them. ... So, it was probably Yahoo!, OCA, Google, Microsoft. I might have swapped around those last two. I can't remember. Google – Google and Microsoft were real close in time. So... but I mean, we didn't really care. We were promiscuous (P3).

However, as a second UC librarian I spoke with explained, this institutional promiscuity regarding digitization, once extended to Google, caused a significant rift within the OCA's leadership ranks:

[Partnering with Google] was seen as a dramatic reversal of [UC's] declared intent, priorities, and motivations. It was also a great surprise to Brewster. I think Brewster felt very aggrieved by this. Felt that the decision made by... primarily Dan Greenstein²⁰⁶... had proverbially left him at the altar, that UC's commitment to the OCA was bound to be enervated by the engagement with Google... um... that, um... that there would be a falloff in activity... and I think Brewster really tried to get UC to recommit to OCA. Which UC did to some extent, um, re-commit publicly to... to both endeavors. But I think the perception that the university would find it all too easy to scale down engagement with OCA was obviously prescient. And... and there are logistical reasons for that as well. It's very difficult to run scanning operations for two different enterprises out of the same facility, um, and so, to some extent, OCA activities ended up being, um, ah redlined if you will – segregated – um, to their detriment (P2).

And in fact, another participant reported that the ill feelings between Kahle and Greenstein ran so deep that the two men essentially hadn't spoken since all this occurred – nearly five

²⁰⁶ In 2005-2006, when all this was occurring, Greenstein was the University Librarian for the California Digital Library, which is responsible for digital library initiatives for the entire 10-campus UC system. He has since moved on from that position (see, e.g., his LinkedIn profile: <http://www.linkedin.com/pub/daniel-greenstein/19/985/473>).

years by the time of the interview (P1).

Despite the scaling back of OCA scanning at the University of California, however, the project's development continued to look promising at least through the beginning of 2008. Five more OCA scanning centers were opened in that year with funding from Microsoft, and still more partners were added (P5, "TRLN Libraries" 2008, Kahle 2008). But starting in late spring of that year, things began to fall apart. First, on May 23, Microsoft suddenly announced that it was bowing out of book scanning in general, having spent \$10 million on its efforts, in order to "focus on verticals with high commercial intent, such as travel;" in essence, they seem to have discovered that book scanning was not, after all, an especially high profit-margin enterprise, despite Google's interest in it (Albanese 2008d, Nadella 2008). At that point, Microsoft released all rights to the scans it had made for Live Search Books, and donated 300,000 such scans to the OCA, helping to nudge that initiative's numbers over the million-volume mark by January 2009 (O'Leary 2009). Microsoft's departure left a major gap in OCA funding which would never be adequately refilled (P5, Albanese 2008d, Guess 2008, Kahle 2008).

Then, later that summer, the OCA's leadership search also ended in failure. Throughout the early months of 2008, the OCA had been searching for a new leader to replace Rick Prelinger, and believed they had found one in Maura Marx, then of the Boston Public Library. Marx was accordingly hired in August 2008. However, after leaving her position at the BPL, Marx began to sense that, contrary to her understanding of the position's terms, Kahle seemed to expect that she would be working for him under the umbrella of the Internet Archive, rather than leading the OCA as a fully independent, collaborative organization. And so she never began the job. Instead, she remained in Boston and founded a separate initiative called the Open Knowledge Commons, and within a few years, became one of the primary individuals driving the establishment of the Digital Public Library of America (P1, P11, "People" 2008, Berry 2009). The Sloan Foundation shifted its digitization funding from the IA to these initiatives (P1), and no other OCA director was ever hired (or even sought).

Lacking both Microsoft's deep pockets and its own dedicated leadership, the OCA quickly withered – or at least lost whatever marginal level of cohesion it had once had. As noted, the final mention of the term "Open Content Alliance" on the organization's own blog appeared in July 2009, and the blog went quiet entirely after May 2010 (Kahle 2009b, Kahle 2010d). Still, many of the activities begun under the banner of the OCA continued; they just weren't called "OCA" anymore. The Boston Public Library, for example, along with other nearby institutions, continued to develop patron-oriented digitization services like scan-on-demand for public domain works and digital interlibrary loan, using the Scribe scanning stations installed at BPL under the OCA ("Have a Hand in Scan-on-Demand" 2008, Colford 2008). Indeed, those scanning stations – which are located in a publicly visible part of the

Copley Square building – still appeared to be seeing heavy use as of my own last visit to the BPL in June of 2011. And of course the Internet Archive has continued to scan books with a wide variety of partners, having established relationships with many content holders during the period of the OCA, as well as many more since (P1, P2, P5, P7, Kahle 2011, Internet Archive n.d.-b). The Open Library continues to improve its services and collections, with digital lending emerging as a particular focus; in fact, in 2011, the IA/Open Library digital lending program won the unanimous support of the state librarians of all 50 US states (Rapp 2010, Kelley 2011, Rapp 2011a). And more recently, the IA has moved into collecting paper books, using climate-controlled shipping containers for storage. Launched with an initial collection of 450,000 volumes in 2011, IA seeks to build its physical book collection both through donations and by reclaiming items deaccessioned by other libraries (P5, Rapp 2011b).

Over the course of its effective lifespan, the OCA facilitated the high-quality, full-color digitization of over 1 million books – a substantial accomplishment on its own – as well as the formation of a number of digitization partnerships that persist to this day. However, ultimately the OCA itself failed to cohere. The multiple online indices to its digitized content – promised by the OCA itself, by Yahoo!, and by Microsoft – all either failed to ever materialize or failed to persist once built (e.g., "Consortium Forms OCA" 2005, "Microsoft Partners with Yahoo" 2005, Crawford 2007). The Internet Archive continues to provide access to the content both through its Ebooks and Texts Archive and through its (much more appealing) Open Library portal, and millions more scans have been added to both of these sites since the OCA *per se* fell by the wayside. But for several of the individuals interviewed for this research, the mention of the Open Content Alliance itself evoked heavy sighs and a sense of regret (especially P1, P2, P11). It had so much promise, and seemed like it could have been so much more, but instead, those responsible for it dropped the baton, and the baton split into pieces. Some of those pieces live on in the Digital Public Library of America; some in the HathiTrust; and many in the Internet Archive and its Open Library project. However, just as the OCA was both less and more than Google Books, so are all of these projects both less and more than the OCA. The balance of this chapter will strive to make sense of the scattered pieces the OCA left behind, along with a few of the traces of it that have continued forward.²⁰⁷

2. Motivations

The motivations behind the Open Content Alliance, as for the Google Books Library Project, were relatively diffuse. Each participating institution chose to join for its own specific

²⁰⁷ And scattered as they are, I hope the reader will forgive some slight oddness in verb tenses throughout. There is real epistemological conflict within this case over what pieces existed in the past vs. exist today vs. never existed at all, as I hope the introduction made clear.

constellation of reasons, and those reasons varied broadly, especially across the boundaries between libraries vs. for-profit tech companies vs. the nonprofit Internet Archive. Still, there are three basic through-lines that can be drawn among these motivations, each of which will be discussed in turn below: (1) the charisma and vision of the OCA's principal figure, Brewster Kahle; (2) the OCA's expressed principles, especially regarding openness, transparency, and preservation; and (3) the complementary yet distinct set of practical objectives prevailing among the different sorts of participating institutions.

2.1 "Brewster, Brewster, Brewster"

It would be impossible to fully understand the motivations behind the OCA without at least briefly exploring the motivations of its main champion and leader, Brewster Kahle. As noted, Kahle was unanimously cited as the most influential person on the project by all of those interviewed about the OCA for this research, and regarding issues of motivation, it is often difficult to clearly discern where Kahle leaves off and the OCA begins: with Kahle constantly at the core, driving the project forward, the OCA became thoroughly infused with his ideologies and aims. From Kahle's own statements, as well as the assessments of interview participants and press accounts, three facets of Kahle's persona – the collector, the tech entrepreneur, and the visionary – emerge as especially useful keys to understanding both Kahle's interest in the OCA and the ways in which Kahle's influence shaped the OCA as a broader project.

2.1.1 The Collector

The concept of Brewster Kahle as a collector was actually first suggested to me by one of my interview participants, a librarian who had worked closely with Kahle in the early years of the OCA. He explained that he thought of Kahle as a collector like any other collector (whether of stamps, cars, or information), with the Internet Archive providing the warehouse for his collections. As he put it,

Brewster is a collector. Have you ever met any collectors? Whether they collect maps or cars, they just... are acquisitive of STUFF. And that's what he is. And collectors don't stop until they've got all the stuff. Um, his stuff is digital. And, um, [...] his collection has no... There's no business model... um... it doesn't need to. Because he's a collector. What's the collector's business model? "Acquire. I buy stuff." It makes a lot of sense when you understand it, at least in that way – I do. He's focusing on the public domain largely. He's had some runs at the in-copyright material, um... there's no... the only rationale you can put behind what he's doing is he's a collector. It explains his resistance to distribution, it explains his resistance to the governance structure, it explains, you know... It's the Internet Archive, the Internet Archive, his collection. So: it's not a bad thing. It's an interesting thing (P3).

This participant went so far as to claim that this was the primary way in which he understood who Kahle was and why he did the things he did, up to and including being fairly indiscriminate about what the IA collection actually contained, and being reluctant to

define the IA's relationships with contributors to the collection beyond expressions of shared enthusiasm for scanning. As this individual continued, "[Kahle] was just 'Come on! Let's just go scan some stuff!' and there wasn't really much um, thought of contract or MOUs, or you know, the usual stuff that define people's relationships" (P3). While Kahle had (and still has) boundless enthusiasm for acquiring ever-more information for his collection, the details of exactly what he was acquiring and how those transfers would actually function seem to have been far less important to him than the simple fact of acquisition.

Other accounts, moreover, tend to confirm that this "collector" mentality is a significant piece of Kahle's persona, especially vis-à-vis digitization, though in less explicit terms. In fact, Kahle himself has made several statements that reflect that mentality to varying degrees. For example, he told the *New Statesman* in 2005 that "The problem with the internet is not that there's too much information out there, it's that there's not enough good stuff. We're looking to radically increase the amount of material on the web" – and in fact, he went on to suggest looking to the "ancient library of Alexandria" as a model,²⁰⁸ averring that digitization presented an opportunity to gather together the majority of the world's knowledge once again, "but then to one-up the ancients by making it available universally" (Hogge). In other statements, particularly regarding the growing IA digital text collection and its recent venture into physical book archiving, Kahle consistently focuses on almost exclusively the *quantity* of information being gathered, to the exclusion of any discussion of the actual *informational content* of the resulting collection (e.g., Kahle 2009a, Kahle 2010c, Kahle 2011, Rapp 2011b). This impression is crystallized in the comments of another OCA participant – also a longtime collaborator of Kahle's – who mused,

Brewster, you know, his metric of success is bits in and bits out. How many bits can we create and ingest, and how many can we redistribute. That's literally all he cares about. He wasn't interested in organizations. He wasn't interested in partnerships, per se. He just wanted to be able to scan (P1).

As will be made clear in the balance of this chapter – and especially in the discussion of the OCA's collections, below – this "collector" element of Kahle's persona and motivations had a powerful influence on the way the OCA and its successor projects have unfolded.

2.1.2 The Tech Entrepreneur

Before Kahle ever began to collect anything, he was an MIT-trained computer scientist and engineer, and more than that, he was an inventor and entrepreneur – and he has continued to be all of those things. He has a keen grasp of the state of the art in digital technology, what might be achievable using that technology, and, often, what might be achievable if he were to build out the next phase of a given technological system (consider his history with

²⁰⁸ As he does with great frequency (e.g., Auchard 2005, Benson 2005, Hogge 2005, Bengtson 2006, Young 2006a, Kahle 2011, Vaidhyathan and Kahle 2011).

data storage systems, for example (e.g., Schiff 2006)). And this facet of his personality also has significant implications for his approach to book digitization, especially in terms of his concept of what is possible and his sense of how quickly things should move.

Both in person and in public statements, Kahle can seem boundlessly, even unrealistically optimistic – and to an extent, he may well be – but his optimism makes more sense when considered in the context of his history with technology. As one individual close to him noted in an interview,

Brewster's an engineer, and when he says that [universal access to knowledge is possible], he...what he is, is saying, is, and I think he'll amplify this when you listen to him talk, ...when he makes his speeches, he's saying it's technically possible to do this (P1).

That is, setting aside all logistical, organizational, and legal issues, technology does already exist with the capacity to make all the information ever recorded universally available worldwide. And in purely technological terms, the point is valid: technologically speaking, a much broader scale of information access is possible than now exists, and as technology improves, so does the opportunity for openness. Still, this narrow focus on what is technologically possible requires setting aside some extremely significant obstacles (again – logistics, organizational dynamics, the law...) – and the OCA ended up facing more than one of those obstacles on its path to atrophy.

Still, perhaps more informative for this case is Kahle's entrepreneurial sense of pacing. More than one OCA librarian noted that Kahle seemed to quickly grow frustrated with the slow pace of action in the library world, and especially with the ponderous process of consensus-building:

to Brewster's credit, he's an entrepreneur, he's made a lot of money, he knows how to start companies. He's gotten frustrated by the way that librarians move so slowly, and just can't seem to make decisions in the time that he probably, rightly so most of the time, thinks that the decision should be made. So I think he was afraid to go down that path, knowing as all of us know, that governance by consensus is a very murky and difficult thing to make work (P11).

Echoing this point, another librarian observes, "I think he comes from the IT world, even though he calls himself a librarian. And I think the IT world is much more comfortable with this whole lack-of-structure-structure." And she goes on,

I think this was one of the difficult parts: when library people got too much involved, they wanted to kind of organize it too much. I guess RLG was involved at one point before it merged, or got together with, OCLC. That was sort of in that final year. And they wanted to manage it too. And I think Brewster thought originally that was a good idea, but basically all of these bureaucratic organizations that we have in libraries just were... just too slow, and too much talk and no action, basically (P14).

Within the OCA, Kahle seems to have been torn between a desire to adopt an open, discursive, democratic approach to scanning, and a parallel desire for that scanning to get moving at tech-company pace, without getting too bogged down in the details of governance. One has the impression that in his prior ventures, he had been able to just make things happen, or rapidly gather others together to make things happen; as such, the slow-moving world of library leadership and consortial arrangements²⁰⁹ seems to have come as a rude shock to him in this regard.

2.1.3 The Open Access Visionary

Finally, one would be remiss not to note the central importance of Kahle's fundamental belief system, centered on openness in all its many forms, to spurring the OCA into existence. Of all of the individuals involved in all of the four cases described in this dissertation, Kahle is by far the one most frequently referred to as a "visionary" (e.g., P5, Roush 2005, Bengtson 2006). Jonathan Bengtson, a librarian who worked with the OCA at the University of Toronto, even contends – and I agree – that "Kahle's vision remains the key to understanding the OCA's goals: first and foremost, to provide free access to the world's written heritage via the web, or at least that part of the heritage in public domain" (2006). It is a powerful and compelling vision, and yet, it is not merely the vision itself that is important to understanding the OCA; it is also the form of the message and the charisma of its messenger. As one interviewee from the IA explains,

I think it's, the vision is very clear. Um, I think we work with a very charismatic leader and founder and visionary, Brewster. So that when you've drunk the Kool Aid, so to speak, here, it's very easy to keep people independently motivated to continue on the path they are, as opposed to "what am I doing here again today? Why I am here?" (P5).

And another, also from the IA, adds,

I mean it's kind of incredible. And most of it comes back to Brewster's... um... insatiable drive to be... um... to make sure that culture stays free. And, I mean, that's one of the reasons I like working here is because the mission is just so obvious (P7).

In essence, Kahle positions himself as an evangelist of openness – and also as the damn-the-man underdog of the tech world – and many highly skilled and passionate individuals follow him gladly, whether through the IA, through one of his myriad other projects, or, while it lasted, through the OCA. And it is this element of his persona that permeates his soaring rhetoric, his inspiring calls to action, and his heated condemnations of Google and its allies (e.g., Hogge 2005, Bengtson 2006, Albanese 2007d, Goth 2007). Kahle-as-visionary is Kahle-as-crusader, and it is in this guise that he gathers the most followers, including and perhaps especially among the librarians involved in OCA. The specific principles that comprise this vision will be described next.

²⁰⁹ A world which brings the phrase "cat herding" readily to mind, along with its associated metaphors.

2.2 Principles

Intertwined with Kahle's persona, the OCA's principles played a central role in motivating participation in the project, especially by the libraries who contributed content for scanning. Several of the OCA participants I interviewed explicitly cited the project's (and Kahle's) principles as a key factor – or even the primary factor – in motivating their institutions to participate in the OCA. For example, an interviewee from the Boston Public Library explains:

I think all of us were motivated by the same principles, which were you know, how do we create an alternative to a commercial thing. How do we really advance this goal of universal access to knowledge? And... My ideas were that would probably be much more of a collaborative process (P11).

And another, from the University of California, links those principles closely to Kahle himself, and to the force of Kahle's personality:

it seemed to me that... um... that Brewster was being driven by noble intent. That he wanted to create as broad of access as possible, um... That he wanted to initiate, um, clear alternative vision. Brewster is a very... ah... large personality, and um... But I do think that, that Brewster was driven by, um, by noble aims (P2).

And finally, a Toronto librarian restates the link between the OCA's principles and Kahle's, and adds that these ideas also meshed well with the values already prevalent in the Canadian library digitization environment:

So, um... but it also, his philosophy – this is Brewster's philosophy – matched with the Canadian environment, which was... all of our institutions are publicly funded. And all of our government grants, which there were a few at the time for digitizing, particularly Canadian content, to put on the Internet. And they were all conditioned on being completely open. So any money that we got, was based on open access. And, so... originally people thought [about Google] "oh well they're going to make a lot of money out of selling digital books" but that really was sort of a fantasy anyway. But the whole Canadian environment was about, more about sharing... because we didn't have as much money, basically, and so the money that governments put in was based on open access. And the same for universities. So we had a kind of similar philosophy and so the Open Content Alliance principles attracted Canadians in general and we have done a lot of collaborative projects (P14).

And while this emphasis on principles is not unique to the OCA – the library partners in the Google Books Library Project also cite principles (and, in particular, concordance with mission) as a motivation for participating in that initiative – the ideological foundations of the OCA seem to have been more strongly and consistently felt among the OCA's participants than the GBLP's were among its contributors. Throughout the available documentation, it is very clear that the OCA's principles – real, purported, or perceived – were a driving force behind its success in gaining support from various quarters.

So what were these principles? Tracing threads through the primary-source accounts, four

dominant themes emerge – three principles and one meta-principle. That is, the principles of *openness*, *transparency*, and *preservation* all ring through strongly, and all alongside and blurring with the meta-principle of what I will call “*anti-Googlism*.” Although this last theme would not be explicitly found in any of the OCA’s official lists of founding principles (e.g., “A Call to Participate in the Open Content Alliance”), it arguably underlies every one of the items that do appear. That is, for each of the OCA’s explicit principles, at some level, there was a substrate of competitiveness with or opposition to Google Books. So the OCA did not simply strive to be open: it strove to be *more open than Google Books*. And it did not just want to be transparent, but *more transparent than Google Books*. And it designed its procedures to be not merely good at preservation, but *better at preservation than Google Books*. The OCA was, as Walt Crawford suggested at the time, both “less and more than a ‘challenge to Google’” (2005) – and yet, the challenge to Google remained ever-present all the same. This anti-Google substrate will be discussed first below, followed by explications of each of the three more idealistic principles in turn.

2.2.1 Anti-Googlism

When the Google project was announced, it made many librarians and others a little queasy. Here were these almost deified centers of learning and knowledge, apparently ceding control their collections to a company that was rapidly becoming the new tech behemoth, in the tradition of IBM and Microsoft before it.²¹⁰ Kahle built upon this unease – as well as the publicity surrounding Google’s project – to expand the IA’s existing digitization operations further in the direction of library books. Specifically, he pitched the OCA as a principled alternative to Google Books: it was what *real* cultural institutions, with *real* commitments to universal, long-term information access would do. For example, in 2006, Kahle averred in an *InformationWeek* piece about the OCA’s progress that, “Google has made a full-court press toward privatizing every library they can get a hold of... But this is a step toward showing there’s an alternative path” (Gonsalves 2006). And by the time of my interview with him in 2011, his tune had scarcely changed: in that conversation, he described the OCA as “*a mechanism of just establishing a set of principles and a set of contributors that wanted something different from the newly announced Google project, which was sort of lock down commercialize the library system... project*” (Kahle 2011).

Still, although Kahle’s melody remained largely consistent, its dynamic shifted a great deal over the course of the project. Specifically, it crescendoed. Early on in the OCA, there was some hope, openly stated by project leaders, that Google would itself sign on to the Alliance; that it would eschew its typical secrecy and find some way to work collaboratively

²¹⁰ As is likely clear by this point, I do not buy into this line of reasoning; however, it was certainly a common line of rhetoric surrounding the OCA-vs-Google-Books divide in the early years of both projects, and it continues to crop up here and there in statements about Google, especially by Kahle (e.g. Kahle 2011, Vaidhyanathan and Kahle 2011).

with those in favor of open content (e.g., Auchard 2005, Crawford 2005). Indeed, Kahle himself expressed this hope soon after announcing the project, telling the *New York Times*,

The thing I want to have happen out of all this is have Google join in....I know we're dealing with archcompetitors, but if there's room for these guys to bend, by the time my kid goes to college, we could have a library system that is just astonishing (as quoted in Hafner 2005).

Of course, it soon became clear that Google was unlikely to sign on to the OCA, and as the project proceeded, Kahle's drumbeat of Google criticism grew steadily louder and more strident. By 2007, his rhetoric about the company had begun to consistently position it – literally – as an Orwellian information overlord, hungry to absorb and privatize not only the entire library world, but the entire information ecosystem. For example, that year he told *IEEE Distributed Systems Online*:

We're talking about how people think, how people pass information on to their young, how we as a society conceptualize ourselves. And if we have that dominated by one or two large corporations, we will be living in an Orwellian world. So the stakes are very high, beyond the financials of it.

...

We have people who are trying to monopolize, to close, the content layer. ... If we've spent the last 30 years building an open infrastructure and we lose it at the content layer, it will all be for naught (Goth 2007).

In some cases Google's name was mentioned and in some cases not, but in every instance it is clear that Google was the intended target. And with a steady stream of statements like these, invariably made in response to requests for statements about the Open Content Alliance, Kahle contributed a great deal to the growing sense in that period that the OCA was little more than an ideological challenge to Google Books. Constantly depicted by its primary spokesperson as an inverse or alternative to that larger project, the OCA grew increasingly negatively defined.

In fact, the growing emphasis on Google in the OCA's rhetoric and statements overall is nowhere clearer than on the OCA's blog.²¹¹ That blog was in operation for about three and a half years, from December 2006 to May 2010. Tracing the term usage across all 49 posts made during that time (in titles and text only, not in tags or URLs), a few trends are clearly visible (as shown in Figure 55). Most salient here is the fact that mentions of "Google" began to exceed mentions of all the OCA-related terms combined ("Open Content Alliance," "OCA," "Internet Archive," and "Open Library") as of 2008, and skyrocketed to nearly triple the mentions of OCA-related terms in 2009 (coinciding with the height of public debate over the Google project's proposed Settlement Agreement). Beyond a certain point, the blog ceased to be primarily about advocating for the OCA, and began to be primarily about opposing Google.²¹²

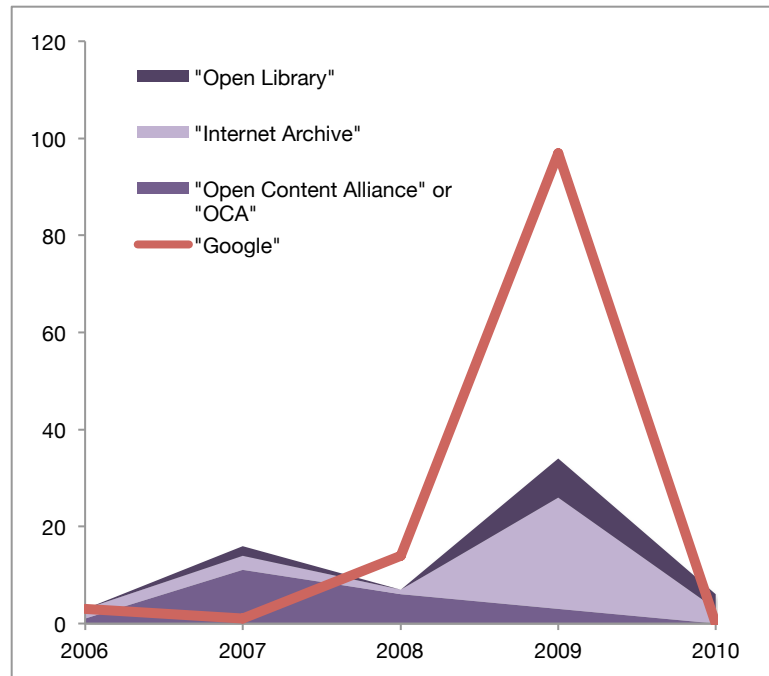


Figure 55: Frequency of Term Usage in OCA Blog Posts, 2006-2010

Notably, the OCA blog was written by a number of different individuals throughout its lifespan: Brewster Kahle, Peter Brantley, Mary Murrell, Karen Coyle, and Linda Frueh. And while many of the more Google-centric later posts were written by Kahle, not all were. Considered alongside other developments during this period, like the establishment of a formal advocacy organization to oppose the settlement by some of the same institutions that founded the OCA,²¹³ this tends to indicate that the trend away from pro-OCA and toward anti-Google was not simply a reflection of Kahle's sentiments, but was felt more widely among those involved in the OCA, especially after the announcement of the Google-AG-AAP Settlement Agreement.

Given the significant extent to which advocacy surrounding the OCA throughout its lifespan centered on critiquing Google Books, it is understandable that the image of the

²¹¹ At <http://www.opencontentalliance.org>.

²¹² A second trend observable in this chart is the marked and steady decline in mentions of the OCA itself after 2007, which parallels the decline of the project as chronicled above.

²¹³ That is, the Open Book Alliance, established by three leading OCA member institutions (the Internet Archive, Microsoft, and Yahoo!), several library and publisher organizations, and Amazon.com ("Members" 2009).

OCA's motivations that emerges does so almost like an outline in silhouette, defined as that-which-Google-was-not. Where Google was closed, the OCA would be open. Where Google was secretive, the OCA would be transparent. Where Google seemed to be proving inept at digital preservation, the OCA would build on a solid foundation of library expertise in that area. Although several of the OCA's ideals would become more contested and/or fall by the wayside in actual practice, both the ideals and the very intentional contrast within each to Google Books remain crucial to understanding how the OCA came into being, and how it flourished to the extent that it did.

2.2.2 Openness

There are two major facets to the concept of openness at play in the Open Content Alliance: *open access* and *open rights*. It will be useful to separate these and discuss each in turn.

Open access is the more straightforward of the two: essentially, it is the idea that all the world's information should be open to everyone, regardless of who or where they are. This concept is embedded in the Internet Archive's official motto, "Universal access to all knowledge," which is emblazoned at the top of the IA homepage and was also spontaneously cited by several interviewees – mostly, though not exclusively, affiliated with IA (P1, P5, P7, P11).²¹⁴ Whether or not those exact words are used, however, the concept of open access, and even universal access, pervades both the interview responses regarding motivations and the public statements made by OCA participants in the early days of the project. For example, one librarian I spoke with noted that "*making the world's knowledge accessible to the world was – is – a collaborative effort*" and that "*Open Content*" was "*not for free in the first place. Not created for free. But, to make it more fairly, openly accessible, is really important. That's what libraries are all about*" (P14). Another discoursed at length about the open access movement more broadly, from the Berlin Declaration to open science journals (P15). And a press account published at the project's outset even quoted Yahoo!'s vice president of search content touching on the virtues of open access to information, averring that the company was "honored to participate in a program that helps further our vision of expanding all human knowledge by working with content creators to make their content available to a growing online audience" (Yahoo 2005). And other such examples abound (e.g., "Consortium Forms OCA" 2005, Kahle, Guren 2006, "Boston Library Consortium, Open Content Alliance Partner" 2007). In this motivation, it should be noted, the participants in the OCA are virtually identical to the library participants in Google Books, who also sought to broaden access to information (as described in section 2.2.1.1 of the previous chapter). Although many of the OCA's participants (and many outside observers) would undoubtedly claim that the OCA was *more* open, or *better* for broadening

²¹⁴ Interestingly, in my interview with Kahle, he noted that this phrase was originally coined by Raj Reddy, the leader of the Million Book Project at Carnegie Mellon.

access, that is a question of implementation. The underlying goal – to expand access – was exactly the same.

Still, there is somewhat more differentiation between the two initiatives' motives with regard to the second aspect of openness – open rights. In this context, the idea of “open rights” can be understood as the absence of additional restrictions on the flow of digital content beyond those already imposed by copyright law. This issue seems to have become a compelling motivation behind the OCA explicitly due to the widespread dissatisfaction with the terms of the publicly-available Google contracts among libraries and other organizations, and especially with the restrictions those contracts seemed to place on what the contributing libraries could do with their copies of the content. As a librarian working with the Biodiversity Heritage Library put it,

I think the Open Content Alliance was truly more appealing is because it is open. And they'll give you anything that they have. Like Google only gave you the files that they wanted to give you. And they didn't give you the rights to everything. We have the rights to everything. Because IT IS OPEN. We said we want to download every single thing that you have in every format that you have, and they said “ok fine, take it!” Google does not allow that to happen (P15).

In his interview, Kahle portrays this contrast in starker terms, seeming frustrated with the Google partner libraries not only for agreeing to participate in that project, but also for then advocating on its behalf:

*so for the out-of-copyright works [scanned by IA], the idea is everybody can have access to all of them. And in bulk. This is in direct contradiction to the Google project. Which is *highly* restrictive, even though anytime you hear libraries that signed up talk about it, you... you'd think they were talking about something completely else (Kahle 2011).*

It is worth taking a moment to tease apart exactly what Kahle is objecting to here. As noted in the previous chapter, the main points of restriction in the Google contracts had to do with limiting automated access to the files, preventing bulk downloading, and generally not competing with the services Google was providing. Also as previously noted, none of the Google participants I spoke with seemed to find these restrictions especially onerous. And in fact, some of the specific “Google restrictions” I heard cited in my interviews with OCA participants – including, to some extent, the claim above, that “Google only gave you the files that they wanted to give you” – find little substantiation in the Google contracts or other accounts of that project.²¹⁵

²¹⁵ All of the Google contracts included the provision of library copies of everything scanned from that library (P12). To my knowledge, the only reason Google ever held back any files was because of legal risk, which led them to move to the escrow strategy for works in copyright later in the project. But that was not so much a matter of Google not “wanting” to hand over the files as it was a function of having scanned in-copyright works in the face of an ongoing class-action lawsuit over that very activity. Of course, it is possible that the participant was noting that with the OCA, libraries could download all the scans, regardless of where the items had been scanned from, whereas Google only

Whatever its accuracy, however, the perception that Google was offering their library partners something radically less than full rights to their contributions, and the lack of transparency surrounding the precise limitations, especially in the early period of the project, created an atmosphere of suspicion around Google Books that never fully dissipated. And although Kahle's frequent claim that Google was "privatizing" library collections seems wildly hyperbolic given the actual circumstances,²¹⁶ it made a compelling battle cry. By contrast, the OCA's promise to place no restrictions on what its contributors could do with their content looked quite empowering and library-friendly.

2.2.3 Procedural Transparency

Another aspect of the Google Books project that disturbed many observers in the library world (and beyond) was its secrecy – and this issue provided another point on which Kahle sought to distinguish the OCA on principle. As previously noted, by 2004, Kahle had spent most of his career building open hardware, open software, and open content systems, in keeping with the sense that the Internet and Internet culture were and should be about sharing all of these things openly at every level. As a close, early observer of Kahle's struggles with the issue of openness vis-à-vis Google noted,

the notion in Internet culture, of secret contracts, sometimes with public entities and um, librarians under non-disclosure, and people whose uh, people in academia who are supposed to be focused on the open exchange of, of scholarship, information, and ideas, suddenly developing these, um, uh, y'know essentially one...opaque relationships that was cause for a lot of people's concern, and if you wanna walk the cat back, probably, um, i- had that secrecy never happened, um, the whole history of this thing might have

gave libraries the items from their libraries – and that much is true. However, in both HathiTrust and DPLA, we are beginning to see the Google scans reaggregate outside of Google, which seems a promising counterweight to that fragmentation.

²¹⁶ Walt Crawford has critiqued Kahle's (and Siva Vaidhyanathan's and Karen Coyle's) use of the term "privatization" to describe what Google Books is doing in a way that I find compelling. As he writes:

I'm not in love with Google by any means. I think OCA is a great idea (although I wonder where the "alliance" has gone, given Yahoo's almost-total silence and Microsoft's diverging effort).

But "privatizing the library system" or, which I've also read, "privatizing the public domain" – I'm sorry, but horsepucky. If Google negotiated exclusive contracts, maybe.

Otherwise, that language is like saying that, if I check a book out from my library that happens to be in the public domain, scan it, and return it to the library, I've "privatized" the book.

Google is borrowing books from libraries (in large quantities thanks to special arrangements), scanning those books, and returning them to the libraries with the promise that the books won't be damaged. Its deals are nonexclusive. Google's scan does not in any way modify the terms under which the book itself can be used.

Google Book Search absolutely expands findability for books and in no way restricts anyone else from building and maintaining book-search systems. Google Book Search for public domain absolutely expands access to the text within books, and in no way restricts anyone else from providing similar access. (For that matter, Google's silly first-page "conditions" are suggestions for use of their PDFs, not legal restrictions.)

How can expansion be viewed as contraction? How can improved access be regarded as privatization?

Want to attack Google? Fine. But is it necessary to debase the English language to do so? Or does it just make a great soundbite? (Crawford 2009).

changed to some degree. If it had been done in a more transparent way (P1).

Kahle was deeply disturbed by Google's opacity, and explicitly positioned the OCA as a transparent, "open" alternative. And throughout its duration, this is consistently the way that the OCA was publicly depicted, particularly with regard to scanning technologies and partnership agreements (e.g., Coyle 2006, Albanese 2007d, Johnson 2007, Kaufman and Ubois 2007, Peek 2007, Ashmore and Grogg 2008). For example, a 2008 article enthused:

Unlike Google's super secret scanning warehouses, OCA scanning centers are located and run by some of its larger partners at eight locations: San Francisco; Los Angeles; Urbana, Illinois; Toronto; Boston; New York; Washington; and London. "They are inside library facilities and those libraries have made the facilities available for anybody to come scan books as long as they cover the incremental costs," says Kahle (Ashmore and Grogg).

And, in some ways, the OCA *was* much more transparent than Google Books: to my knowledge, no non-disclosure agreements were ever signed, or even requested, as part of the OCA,²¹⁷ the scanning centers were (and some still are) open to examination by anyone interested in looking at or learning about them,²¹⁸ and those involved have generally seemed quite willing to discuss the procedural details of the OCA's scanning with all and sundry – all in radical contrast to Google's practices in the same areas. Although there is an important discussion to be had about transparency at an interpersonal level vs. transparency at an active, informing-the-public level (to which I will return in section 4.1.4 of this chapter), the OCA did at least manage to avoid some of the points of non-transparency that so troubled librarians about Google Books, and thereby remained basically acceptable in that regard from the perspective of its actual and potential library partners.

2.2.4 Preservation

There was a strong sense among the OCA's participating librarians that the Internet Archive was in sync with their wishes and needs when it came to preservation issues, to a much greater extent than Google was: as one OCA partner librarian explains, "*to the extent that we're focused on long-term preservation, I think that's where we have very much a common ground, and I think Brewster has some good ideas*" (P14). Specifically, there seemed to be a sense that the IA, as another memory institution, had actually thought about preservation issues, and had reasoned processes for dealing with them, especially in terms of image quality and collection redundancy. During the interviews, one librarian told me flat out, "*You know the Internet Archive actually had a much better quality of scanning*" (P15), and my own experiences using IA-scanned books tend to affirm this estimation – as does the fact that unlike Google,

²¹⁷ Whereas at Google, one must sign an NDA simply to walk through the door, and they have lawyers on call to talk you through it if you resist (as I did).

²¹⁸ In fact, the scanning center at the BPL has one wall open onto a public hallway in the central library building.

OCA was never substantially criticized on this point, even in pieces discussing such issues in both projects (e.g., Coyle 2006, Crawford 2007, Goldsborough 2008, Leetaru). The ability of the Internet Archive to produce images whose quality more closely approached prevailing library preservation standards was an undeniable selling point for the OCA. This too will be revisited below.

Still, there was a second preservation principle which seems to have been even more compelling in drawing library support for the OCA: that is, collection redundancy – and in particular the extremely broad-reaching redundancy that can be achieved through full openness. In the interviews, one librarian describes how she came around from being nervous about opening access to her library's collections to seeing the merits of that approach as a preservation strategy:

Preserving heritage and information in the digital world is just... it's fragile, but it's fascinating, and will it work if enough people have it everywhere? I was more worried about it at the beginning than I am now, because I see how many people have copied what we've done. Some of them have done really poor copies. Some of them are re-selling them in awful fashion. But on the other hand, it's... it doesn't seem possible to easily squash it now. Even for ones that we think we should possible take down because of their mistake, or something, they're just everywhere. So you can't really eliminate it. Nobody controls it (P14).

Essentially, the idea here is that if enough people have enough copies of something, it's hard to make that thing disappear – and making books openly available on the Internet virtually guarantees that they will see wider and more redundant distribution than they would in an equivalent proprietary system. This piece of the OCA truly takes the preservation aims of the Google Books participants to another level. In Google Books, redundancy was praised as a strategy, but in practice was only allowed to spread as far as a limited number of approved server locations, at least with regard to rehosting large portions of the collection (e.g. Google servers, library servers, HathiTrust) – and as such, the removal of content from those few servers could effectively make it disappear. In the OCA, however, the door was wide open: for many items, anyone who wanted to download, aggregate, reuse, rehost, or redistribute them was free to do so in any way they liked. As such, rather than having a limited set of central control points, the OCA was meant to operate more like the Internet itself, where the removal of any one node (or in this case, any one digital book copy) would leave the overall network (or collection) unharmed.

2.3 Pragmatics

The pragmatic, utilitarian concerns that motivated participants in the OCA were quite similar to those detailed for Google Books; as such, they will be covered in less detail in this chapter than in the previous one. Indeed, one participant involved in both initiatives described the complementarity of practical interests among the collaborators as a key

similarity between the two digitization initiatives:

I mean any of these consortial activities are partnerships, and partnerships work best when the partners have complementary interests, that they're not the same... that it's not everyone trying to pursue the same thing. And I think that's true in every case. And I think that if you looked at each of the partners, you'd see that their interests were different, right? So, Yahoo! was interested in quick, free, or cheap marketing. And they got it. Um OCA... You know, Brewster was interested in building his collection, you know, extending into new areas for him, which was digitization of... books. ... You know, and we had a collection management sort of thing. And then if you went around the guy from the Smithsonian, Tom Garnett, you know, he'd been trying to do things, y'know, to expose his collection for a long time, and if you went and looked at the Google Five you'd find that Harvard and Michigan and Stanford and the others all had their own very different reasons, um, uh, but, you know, and, as did Google, but they were all served well in that consortium. And... you know [the OCA and Google Books are] both good examples of partnerships at work. Because of that complementary interests of the partners (P3).

Kahle's charisma as a leader and the principles espoused by the project played relatively uniform roles in motivating participation in the OCA by a wide variety of institutions. However, alongside those more consistent motivations lay a mosaic of more differentiated ones, specific to individual institutions or subgroups thereof. And as the passage above argues, this mosaic of interests can be a virtuous thing, when it comes together to form a mutually beneficial pattern of collaboration. And so it did, for a time, in the Open Content Alliance. Three such complementary interests – the desire of the tech companies involved to improve their respective technologies, the interest of libraries in modernizing and improving their existing functions, and the promise of reducing the costs of digitization, also generally within libraries – are briefly described below.

2.3.1 Improve Technologies

Unlike Google, the technology companies involved with the Open Content Alliance expressed uniformly business-related motivations for their participation. In an early IA-Yahoo joint press release, for example, HP and Adobe explicitly cite their existing leadership in digital imaging, and their intention to build upon that expertise to develop efficient workflows and rich user experiences surrounding digital books within the OCA (Yahoo 2005). That said, beyond this press release, neither of these companies had very much – if anything – more to say about their involvement in the OCA, so it is unclear to what extent they actually remained involved in these areas. But to the extent that they were involved, they seem to have been motivated by the potential to extend their core business in digital imaging.

More clear and more publicly communicative, though no less business-oriented, was Microsoft. Throughout their involvement in the project, Microsoft was very explicit about the extent to which its participation was motivated by a desire to improve its search

systems (at the time, grouped under the banner of MSN Live Search). In announcing its participation, the company asserts:

MSN Book Search will help address the fact that over 50 percent of people's online queries go unanswered today on search engines, according to internal Microsoft® research. This effort will enrich people's search experiences, allowing them to access and interact with previously unavailable digitized data in contextually relevant ways, such as facilitating book club interactions or informal family reading, indexing images, or using graphics, keywords and other features to allow for better and faster retrieval of information (MSN 2005).

Essentially, Microsoft wanted to scan books in order to put more high-quality content within reach of search engines – a motivation very similar to some of those expressed by Google (See previous chapter, §2.1.3, “Get More Data”). As the project progressed, Microsoft representatives continued to repeat this basic point – that the company's participation was all about providing the most relevant possible search results to users (e.g., “Microsoft Partners with Yahoo” 2005, Young 2005b, Guren 2006, Guren 2007). Additionally, although Yahoo! generally had very little to say about its motivations for joining OCA, the evidence that does exist tends to suggest that they were in it for the search index and functionality improvements as well (Hafner 2005, Yahoo 2005).

A clear undercurrent to both of these two companies' emphases on improving their search technologies, once again, was competition with Google. At the time, Google had already started scanning books in a big way, but they were being very secretive about exactly what they were doing, and how they expected to realize profit from it (if indeed they did). It seems quite plausible that Microsoft and Yahoo! got into book scanning in part to try to figure out what Google saw in it; whether there was some goldmine there that they would miss cashing in on if they didn't jump on board soon (in my estimation, this seems truer of Microsoft than of Yahoo!). Ultimately, as noted, both companies jumped ship fairly quickly. Based on available evidence, it is reasonable to suppose that both of these companies rapidly discovered that there was no money in this – or at least no significant money within any reasonably efficient timeframe²¹⁹ – and thus turned away from book scanning to focus on more immediately profitable areas of their respective businesses (as, indeed, Microsoft's final press release on the topic explicitly states (Nadella 2008)).²²⁰

²¹⁹ After all, recall: book scanning is *extremely* expensive in aggregate, at the kind of massive scale Google was pursuing. And between Google's first-mover advantage and the fact that most of the books that any of these projects were going to be able to make available from library collections were very old and fairly obscure (and thus not likely to be especially profitable, even were they not being given away for free), it seems clear that it would take some time to realize any reasonable return on a significant investment made in library book scanning.

²²⁰ Although Yahoo's reasons for departure were never made publicly explicit, it would not be surprising if its reasoning followed a similar trajectory to Microsoft's.

2.3.2 Keeping Up with Technology (“Modernization”)

As in Google Books, many librarians involved in the OCA cited a desire to keep the library relevant in the digital world as a motivation for participating in large-scale digitization. In particular, the OCA partner librarians seemed excited about the opportunities provided by digital technologies to expand upon their existing missions and services. As one such librarian explained,

I think what OCA really wanted to do was to create an information infrastructure that would bring that spirit of the Boston Public Library, that early spirit, that free to all, that knowledge is our right, and... and a real duty for us to pass on as the basis of democracy. It took that and wanting to update that into the digital age (P11).

And another described, on a more pragmatic level, how digitization could help the Biodiversity Heritage Library better support its central user base:

this is allowing scientists around the world to have access to things that they---you know, a lot of this stuff, to be honest with you, is rare. I mean there aren't that many books on, you know, [unintelligible; “oripathy” maybe?] or something like that. And they're, they're rare because only two or three people worked on them, they were in the great museums of the world, like the American Natural History Museum, or the British Natural History Museum, so these books are rare. And you know you would have to drop everything, fly over to England or fly to South America, to be able to look at them. Now, they're there! There online. They're available for you to examine. And you can examine them in different ways. You can do a lot of data mining by using the indexer and organizer. And you know, I just think it's terrific (P15).

As these passages show, the OCA librarians' sentiments in this area were largely extremely similar to those expressed by Google library partners regarding keeping their libraries and collections relevant and useful to their patrons in the digital world, as reported in the previous chapter (§2.2.2.2). And it is not just the librarians – these views are also echoed by those on the Internet Archive end (P1, P2, P5). Indeed, Kahle himself expounded at the time that in order to reach the dream of “Universal Access to All Knowledge,” the OCA's members were working to “bring our libraries into the digital world by building and offering digital services in the tradition of our institutions” (2007b).²²¹ Although partnering with the Internet Archive might not have had the “cool factor” of partnering with Google at the time, the drive to keep libraries on pace with modern technologies was equally important in motivating librarians to participate in both projects, and sympathy with libraries in this regard appears to have helped motivate the Internet Archive's leadership in the OCA as well.

²²¹ That last phrase, “in the tradition of our institutions,” moreover, is another fairly clear dig at Google, which was perceived by many critics as a radical departure from library ideals and practices by the institutions who agreed to participate (see especially the work of Siva Vaidhyanathan, e.g., 2007, 2011).

2.3.3 Cost Reduction

As in the Google Books Library Project, librarians and other content providers signed on to scan with the Open Content Alliance in part to reduce the costs of digitizing their collections. Several librarians interviewed cited this quite directly as a motivation (P2, P3, P14, P15). This theme is similar enough to the equivalent one among Google librarians that it merits little further elaboration, except on one point. That is, where in the Google case, librarians were motivated by a desire for Google to *cover* their costs, in the OCA that sort of nearly-full cost coverage was not necessarily on offer – the only exceptions being the libraries whose costs for digitizing particular collections were picked up by Microsoft and/or Yahoo! (e.g., P2, P3, P5). Rather, the OCA, via the Internet Archive, generally offered to *reduce* libraries' digitization costs by setting up and supporting more efficient scanning workflows, paid for by the libraries (or the libraries' donors) themselves. The implementation of these systems (discussed in more detail below) made scanning much cheaper for the libraries than would have been the case if they had continued with their previous digitization strategies, and that reduction in cost was a significant factor in libraries' decisions to participate (e.g., P14, P15).

3. Definitions

The section above describes the diverse motivations expressed among the early collaborators in the OCA for engaging in digitization through the Open Content Alliance. As in previous chapters, this section will explore some of the ways in which these same stakeholders envisioned their project's potential user base and collection scope (in that order), and how those visions relate to the underlying motivations described above.

3.1 Users

The vision of the user in the OCA is quite similar to that for Google Books, but with a few twists. Similar to the Google case, there is a strong sense among participants that the project should be open to *everyone*: the word “universal” comes up frequently in both internal and external references to the project's scope of access, often drawing directly on the IA's institutional motto of “Universal Access to All Knowledge” (e.g., Auchard 2005, Hogge 2005, Bengtson 2006, Ashmore and Grogg 2008, Frueh 2008) – and, as cited above, one librarian interviewed also made a direct connection between the OCA's intended user base and the “free to all” ethos of her home institution, the Boston Public Library (P11). Also similar to the Google case, the content-contributing partners tended to focus more strongly on educational users and their specific constituencies than the Internet Archive did (e.g., P14, P15, P18, University of California Libraries 2005), although perhaps not as vocally as the Google partners. Genealogists are also mentioned as potential users of the OCA, as are independent scholars and personal acquaintances/family – and like in the Google case, the non-librarian project leaders seem more excited about these users than do the librarians

involved (P1, P5, Kahle 2010b). And finally, as for Google Books, it will not be especially useful to dwell on the actual number of estimated users for the OCA, for the same two reasons: lack of data and breadth of scope. One senses that when OCA participants say “everyone,” they mean it, and so the scope of the user base becomes coterminous with the combined user bases of the sites through which access to OCA content can be obtained (more on this below, in the discussion of “Structures”).

Still, as in the Carnegie chapter, there will be more to gain here by highlighting the unique features of the imagined user among OCA participants than those features shared with Google Books. As such, the balance of this section will do exactly that. Specifically, it will explore four intriguing and unique elements of the vision of the user expressed by OCA participants: (1) the active unwillingness to foreordain any kind of user group, in keeping with the principles of openness described above; (2) the issue of literacies – both technological and traditional – which come up more directly here than in GBS; (3) the explicit focus on creating a platform on which others can build (though there is some divergence of opinion within the data as to whether commercial uses would be welcomed); and (4) (related to (3)), the focus on ensuring that the data produced by the scanning remains open to automated extraction and other machine-based uses (which is vastly different from Google).

3.1.1 The Anti-Definition

Far more so than in any of the other three cases examined here, the OCA’s leaders and participants projected an active aversion to pre-specifying any particular user base for the initiative, out of a sense that any such foreordination would tend to bias the services toward those preferred users. This comes through especially clearly in two interviews – one with a librarian from Boston Public, and one with an interface designer for Open Library. The former cited the history of her institution in explaining her position, telling me:

We try not to really think about who’s using things how. That’s... not that every thing is for everyone, but... Um... You know we try not to prescribe use. You never know what the use is going to be. You never know whether it’s going to be... You know, we used to have a thing hanging up in the lab at the Boston Public Library, it was this article about the Gallo Brothers and the way the Gallo Brothers started their wine business was they went into a public library and they were digging around for something and they found a book on winemaking. You never know who is going to use your information for what purpose. So, you know... I hope that everyone can use it (P11).

Essentially, the idea here is that information services should be designed to support serendipity, rather than tracking users into a predefined path of learning or discovery. To borrow – as the participant has – from the history of public libraries, it is somewhat like the philosophy behind the open stack: if you allow patrons to simply wander about, without trying to guide them too much, they will discover and create things you would never have

thought to guide them toward.

These sentiments are elaborated upon with a tech-industry slant by the designer for Open Library, who expressed her aversion to the common practice of using stereotypical user personas (e.g., Cooper 2004) to guide design:

I try not to do that. I try not to have personas in my mind when I design things....Um, because I think the more you focus on a specific persona, and his or her needs, the... you distance yourself from what the software's actually trying to do. Like, you know, the classic kind of "oh she's a soccer mom, she's 40, she doesn't have any time to do stuff, you know, she buys Cosmo, blah blah blah" ... why isn't she as capable of editing something as a librarian? Or why does she need special assistance to do some... to perform some actions in some piece of software? You know I think it's kind of actually... actually can be kind of pejorative sometimes to use personas in software design, because all your own cultural assumptions about this soccer mom are... you know projected onto the software, and... so, I try to say, um, how generically useful can I make the system? So anybody can use it. [...] As soon as you begin tailoring software for a specific group, you... y'know, you divide its function. And that's... I'm a huge fan of trying to make utilitarian software that anybody can use. So I don't have anyone in mind sitting using and use it. Apart from an interested book lover (P7).

As for Google Books, the aim within the OCA was to make the end product usable by as many different sorts of people as possible. And for this participant, responsible for the creation and maintenance of one of the main interfaces to the OCA's output, that meant refusing to stereotype, condescend to, or gerrymander its user base – or even to let any consideration of users' personal characteristics inform the design at all.

In fact, the reluctance to specify or tailor interfaces to particular audiences seems to run even deeper in the OCA than at Google (beyond those already cited, this also came through in interviews with P1, P14, and P15). I would suggest that this is so at least in part because the OCA's studied user-blindness emerges as an extension of principle, whereas Google's similar reluctance was based on a more pragmatic assessment of their existing reach. That is, Google simply *expected*, without really having to think about it, that Google Books would serve "everyone" – or at least billions of people – because that is what all of Google's services already did.²²² There was no need to think about particular kinds of users for Google Books because those leading the project already knew they would be able to tap into the incredible vastness of *Google's* population of users, especially once they began blending Books results into Google web search. Massive global audiences are simply what Google does. For the institutions participating in the OCA, on the other hand, breadth of audience was not "built in" in the same way. The Internet Archive, the BPL, the Biodiversity Heritage Library, the University of Toronto – these institutions couldn't just assume that any

²²² At least, that is what Google's *successful* services do. And if services are not successful – or even if they are simply not adding a great deal of value to the company (see: Google Reader) – Google is hardly shy about shutting them down.

reasonable facsimile of “everyone” might happen across the resources on their respective websites and find them useful. In this kind of situation, one reasonable response might have been for the OCA to try to tailor; to find a niche or a set of niches for their services to sit in (and indeed, some of the content-providing institutions – especially the Biodiversity Heritage Library – have done this). But serving a niche does not fully accomplish the principled aims of openness and inclusiveness upon which the OCA was founded. And so instead, in the OCA, and especially among generalist institutions participating in the OCA (e.g., BPL, IA), one finds this rejection of the very idea of contemplating the user base, on the grounds that any such contemplation might pigeonhole the resulting services, effectively closing them off to unanticipated users and uses and thus failing to achieve the ideal of full inclusion and openness.²²³

Despite this philosophical stance, however, some basic outlines of the OCA’s imagined user do bleed through, even here, and even among those most reluctant to specify. The remaining subsections below will discuss three types of user that emerge as especially favored within the OCA: the (Internet) Literate User, the Maker, and the Machine.

3.1.2 The (Internet) Literate User

Any book digitization initiative would tend to demand some level of literacy on the part of its users; after all, the literate comprise the dominant user group for books themselves. And any such initiative would tend to require some level of technological literacy as well: if the books are on the Internet, one needs to be able to use the Internet in order to get to them. Both of these facts are true of both the Google Books Library Project and the OCA, as well as many other book digitization projects. However, even beyond these basic levels of literacy, the OCA – especially as implemented by the Internet Archive – seems to envision a particularly intellectually engaged and technologically adept type of user. As one IA staffer suggested regarding the IA’s interface for books (archive.org/details/texts),

I might be projecting here, [but] I think it’s more of a confident user and competent user who would come to our site. Somebody who’s willing to engage, test, and try. As opposed to someone who needs to have everything laid out. You know AOL tried to make the web experience easy for the non, the non-user. Um, I think it’s tougher if you’re a non-user. That type of person would be...you can get enjoyment out of it, but you might not

²²³ Now of course, the question of whether using personas or otherwise specifically contemplating a service’s user base actually tends to bias or restrict the service is an open one in interface design. Personas are popular for a reason, not least because without them, many believe that designers simply default to designing for themselves and people like them. (Indeed, instances of this kind of self-as-“everyone” elision in service design have been noted in each of the three cases preceding this one.) And to the extent that this is the case, such designs can be far more exclusionary than designs created using more concrete methods for thinking through the eyes of one’s user base – after all, in the case of digitization initiatives, not everyone is a web programmer, but most interface designers are. Avoiding talking down to your audience is undoubtedly a good thing; it only becomes problematic when doing so leads to the creation of services that that audience lacks the skill set to use.

uncover as much of the value that's there (P5).

In order to get the most out of the IA's interface, that is, one has to be willing to work for it – and that kind of work takes skills.

This theme of putting some work into one's Internet use was echoed by Kahle, who explained,

We think of the main core user for the Internet Archive service as the Wikipedia generation. ... People that are oriented towards using the Internet, um, as an information resource, and not just to either check mail, or you know watch YouTube, you know. But, do some level of participation. Right? So, read. So yeah, we kind of use our... our target market is the Wikipedia generation (Kahle 2011).

He quickly goes on to explain that his use of the term “generation” is not actually meant to imply young people, necessarily, but rather that cadre of Internet users who view the Internet as an interactive space and not merely another form of consumer entertainment. But again, this cadre, by definition, would have to be composed of relatively advanced Internet users – those who not only understand how to get to a website (which not everyone does – far from it (e.g., West 2011, Zickuhr 2013)) but also how to go deeper: how to edit a Wiki, how to run a complex search, how to parse metadata, how to gauge reliability. Second nature as they may be to everyone at the IA, and likely also to most readers of this text, these skills remain far from universal – and the IA's orientation toward those who possess them restricts its audience, despite whatever protestations of non-specificity it might make.

3.1.3 The Maker

The second favored user type for the OCA, the Maker, could be seen as an amplification of the first type, with an element of entrepreneurialism thrown in: an individual (or business) with both the skills and the initiative to take vast tracts of text content and use it to create new services, new interfaces, and new products. Depictions of this type of user emerge consistently throughout both the project's public statements and my interviews (e.g., P1, P5, P7, P15, Suber 2005, Albanese 2007d, Kahle 2007b, Kahle 2011). For example, the project's central FAQ page explicitly states that “the OCA welcomes all efforts to create and offer tools (including finding aids, catalogs, and indexes) that will enhance the usability of the materials in the archive” (Open Content Alliance n.d.-b), and Kahle expands upon this point in a later public statement, averring that the OCA's rights model

is a bold experiment in openness. Like Wikipedia, the data is editable and bulk downloadable. We see this as crucial to build a solid framework that can grow and evolve as well as inviting other non-commercial efforts to leverage the collection.

This means that others can make custom interfaces to the same data, say for different language speakers. This also means that some categories of books can

have different interfaces; telephone directories can be treated differently from novels and differently from travel guides (Kahle 2007b).

The concept of the DIY interface is one that will be returned to in the discussion of structures, below. For the moment it will be sufficient to note that those willing and able to create such additional interfaces were not only a foreseen and desirable user group for the OCA, but also one that those leading the project actively sought to recruit through statements like those above.

Additionally, it is also worth pointing out that the emphasis on this particular type of user was likely amplified in part because it provided yet another clear point of contrast between the OCA and Google Books. To the extent that there was a community of programmers out there, ready and waiting to build new services on top of millions of digital books, their efforts would be welcomed and encouraged by the OCA where they had been thwarted by Google's restrictions on bulk downloading.

3.1.3.1 *Commercial vs. Non-Commercial*

Still, an intriguing tension arises among statements by OCA participants regarding the question of reusing OCA content for commercial purposes. On the one hand, at several points, OCA participants proclaim the virtues of opening up the collection for commercial as well as non-commercial reuse. For example, one librarian cheerfully muses,

if someone takes you know some of the illustrations that you've scanned and then they make a million dollars on them. Well good for them, let's create jobs. We're happy for them, you know? Stupid we didn't think of it ourselves. But it's... ah... you know, I think of this as a job creator (P15).

And Kahle himself indicates at various points that commercial repurposing of OCA content is not only allowable, but desirable; that if "someone wants to print and binding [sic] a book and sell it on Amazon.com – go nuts, if they want to make it into an audio book and post it on the web – go for it (we will even supply the hosting for this)" (Kahle 2005a). And yet, the Open Content Alliance FAQ states that it welcomes "*non-commercial* efforts to leverage the collection" (emphasis added), and the Internet Archive's own Terms of Service page indicates that users must "certify that [their] use of any part of the Archive's Collections will be noncommercial" (Internet Archive 2001). Confusing matters still further, moreover, each contributor to the scanning²²⁴ was actually allowed to place its own set of restrictions on reuse (Kahle 2005a), so discovering what the actual restrictions are on reuse of any given OCA-scanned book is far from straightforward (Leetaru 2008). As such, it remains unclear what reuse restrictions might actually exist on, say, a bulk download of 100,000 OCA books from the Internet Archive. Can they be used to build a commercial service at all? Can some

²²⁴ Often more than one institution for each book, with some partner institutions providing technology or financing and others providing the books themselves.

of them and not others be so used? The available guidance in this area seems murky at best.

3.1.4 The Machine

The final favored user for the Open Content Alliance is not what one might typically think of as a “user” – in fact, it’s not even human. It’s a machine. More specifically, it’s a computer – or, more likely, several. Alongside their relatively vague visions of the content’s human users, OCA leaders frequently cite the benefits of keeping the content open to direct, automated probing by external search engines and data mining tools – again in direct contrast to Google’s policies on this issue (P7, P15, Kahle 2005a, “Internet Archive Receives Grant” 2006, “Boston Library Consortium and Open Content Alliance to Provide Digitized Books” 2007, e.g., Open Content Alliance). One of the pieces of the Google Books Library Project that Kahle and his allies found most objectionable was its prohibition on machine access, and it thus makes sense that in his own digitization initiative, Kahle would strive to keep that barrier down. And for the most part, he did so, even obtaining consent from Yahoo! and Microsoft to allow their direct competitors, including Google, to index content whose digitization they had supported (Young 2006a). Although the same caveats regarding mixed messages within the restrictions apply here as they do above, for the most part the OCA did maintain this commitment to allowing automated access to its scans. The machines remain free to browse.



In a sense, these four visions of the user are all of a piece – and in a sense, they lead to the same end, which is to generally privilege technical accessibility over facilitating effective intellectual access to the materials (in analog terms, one could say this is like placing a book on the shelf, but not checking to see that your users can actually reach that shelf). The project’s leaders, especially at IA, did not want to specify a user base, because that might close off potential uses. But at the same time, when those leaders do talk about users, it is clear that the users they actually have in mind are very internet and technology literate, and possibly capable of using programming to extract the data from the database and build something new on top of it. And that is not “everyone.” This sort of perspective has historically been quite common within computer science and engineering work on information access systems, up to and including digital libraries and LSDIs (see, for example, Adam and Yesha 1996, Adam and Naqvi 1996, Schatz and Chen 1996, Schatz and Chen 1999, President’s Information Technology Advisory Committee 2001, Langley and Bloomberg 2008). From this perspective, providing “access to information” is presented as a relatively simple matter of making the information *technologically possible to get to*. But that is only half the battle in information access – if even half. For actual *intellectual* access to information to occur, technological access is a necessary but insufficient condition; on top of that, the skills of the user must mesh with the design of the system and with the

information it contains. That is, information access requires not just being able to *touch* the information (or being able to make your computer touch it), but also being able to *navigate* and *understand* it. For users lacking the requisite literacies, access will be at best incomplete, and at worst entirely thwarted.

To be fair, of course, to the extent that this perspective creeps in among the participants at the OCA, it seems more a reflection of optimism and faith in humanity than any kind of active desire to create obstacles. Just as Andrew Carnegie's library donations were shaped by Carnegie's boundless faith that most people would be, like himself, able to use libraries unaided for self-improvement, the access systems advocated and produced by the OCA (and especially by the Internet Archive) reflect a great deal of confidence that most people, like those systems' designers, would be able not only to use relatively complex information technologies, but also to construct new and even more advanced systems on their own. Of course, Carnegie's expectations in this regard went unrealized; as discussed in that case chapter, one of the core issues that led to the program's termination was that most library users did, in fact, need guidance, and that there was no point in continuing to donate infrastructure while neglecting the quality of service (Johnson 1916). And despite the many differences between the technological and social contexts of Carnegie's initiative vs. Kahle's, one suspects that this particular lesson may bridge the divide.

3.2 Collections

Moving from users to collections, within the OCA, as in the Google Books Library Project, there is a significant gulf between the public rhetoric of striving for "Universal Access to All Knowledge" and the private assessments of individual OCA leaders regarding the feasibility or even the desirability of pursuing truly comprehensive book scanning through the systems that Alliance put in place. This section, as in the previous chapter, will proceed in three parts, examining first, the quantitative growth of the OCA's digital book collection and that of its successor projects at the Internet Archive; second, the OCA leadership's perceptions of the meaning and possibility of "comprehensiveness" where this project is concerned, and finally, to the extent that such things can be ascertained, some discussion of the composition of the aggregated collection, and where its boundaries might lie.

3.2.1 Size

Where in the Google Books Library Project, the main obstacle to tracking the size of the collection has been corporate secrecy, in the OCA the main barrier is instead the lack of any central, dedicated system for tracking that number. Instead, a multiplicity of related numbers are available – the volume count listed at the top of the IA's Ebook and Texts Archive page (Internet Archive n.d.-b); the publicly-announced milestones in scanning from particular institutions like the Boston Public Library, Cornell University, or the Library of Congress (P11, Crawford 2005, Kahle 2009a, Kahle 2010a); the number of books

scanned by Microsoft and later donated to the OCA (O'Leary 2009); and more – and these numbers all overlap, and in some cases also extend well beyond the OCA itself, to include the work of other digitization programs at the partner institutions as well. In fact, the progress of the OCA *specifically* was only ever announced at four points: three times in 2007 (upon reaching 130,000 volumes in March, 200,000 in October, and 250,000 in December), and then once in January 2009, upon surpassing 1 million volumes (Goth 2007, Hane 2007, Kahle 2007b, Ashmore and Grogg 2008, O'Leary 2009). In addition to these points, two more can be added which, though not strictly limited to the OCA, seem fairly clear reflections of activity within it: that is, the figure of 2800 volumes scanned in the first year of the IA/University of Toronto pilot project (Crawford 2005), and the announcement in December of 2006 that the IA had made 100,000 books available through its Ebook and Texts Archive ("Milestone Achieved"). Neither of these announcements explicitly mention the OCA, but their relationship to it can be inferred from their temporal and institutional contexts.

In fact, this last figure provides a clue as to how one might get a sense of the OCA's contributions in terms of volume count, despite the relative paucity of solid numbers on that initiative specifically. That is, one of the core structural features of the OCA was that all its scans were to be deposited with the Internet Archive for long-term preservation (Open Content Alliance); as such, the IA represents one of the few, if not the only, place on the web where all these scans were actually brought together in aggregate. The part of the Internet Archive dedicated to these types of materials, its Ebook and Texts Archive,²²⁵ moreover, lists a rough count of the items it contains at the top of its front page. And conveniently, it is possible to use the Internet Archive's own Wayback Machine²²⁶ to trace this top-of-page figure back in time, by pulling up archived versions of the Ebook and Texts Archive homepage going back to 2005. This figure will not be completely accurate for the OCA of course; among other things, the IA was already scanning books (and other texts) prior to the start of the OCA, and the IA also allows any registered user to upload additional texts to the site. As such, the IA's Ebook and Texts Archive figure will be consistently higher than the actual count of volumes scanned via the OCA. Still, comparing the cached IA Ebook and Texts Archive volume counts to the few numbers that are known for the OCA, it would seem that at least during the period from 2007-2009, the approximation is actually quite close. Figure 56 illustrates this trend, using figures from the last archived version of the IA Ebook and Texts page in each month from January 2005 to September 2013 (in purple), superimposed with the few OCA figures that are known (in

²²⁵ <http://archive.org/details/texts>

²²⁶ <http://archive.org/web/web.php>

green).²²⁷

Beyond the figures already noted, moreover, a third line appears in Figure 56, dashed, in red: this is a (very) rough estimate of the contributions of Google Books scans to the IA collection by user “tpb” – that is, the Pirate Bay (P8). It is not entirely clear when this user (or group of users using the same screen name) began executing its Robin-Hood-esque transfer of files from Google to IA, but it seems reasonable to place the starting point shortly after Google began to offer PDF downloads of its scans, in late 2006 (this is further supported by a cursory examination of the dates on the “tpb” files present in IA, the earliest of which seem to originate in 2007). Whenever it started, and however it grew in the intervening period (also entirely unknown), the number of books added to IA in this way by the time of this writing had grown to nearly 908,000 – almost twenty percent of the Internet Archive’s entire text collection – and possibly more.²²⁸ This figure has been included here to illustrate one of the many ways that texts can enter the Internet Archive – and thus inflate its volume count – beyond additions from the IA’s own scanning initiatives.

²²⁷ There are two exceptions to this: the first figure is the earliest figure provided for January 2005, which is the first time the site appears to have been archived, and the final figure is the number that was listed on the live page itself on the date on which this data was collected, September 17, 2013. Also, there were two months where the page was not crawled, and thus no data exists, in November of 2009 and August of 2012.

²²⁸ This figure was ascertained by executing an advanced search on the “description” field on IA for the term “tpb.” Many, perhaps most, of the Google scans added to IA in this way bear some version of the sentence, “Book digitized by Google and uploaded to the Internet Archive by user tpb,” but since the exact phrasing varies, the single-term search seemed most accurate. It is, of course, completely possible that not all Pirate Bay uploads from Google bear this description, and so this figure may well be an underestimate – possibly even a very large one.

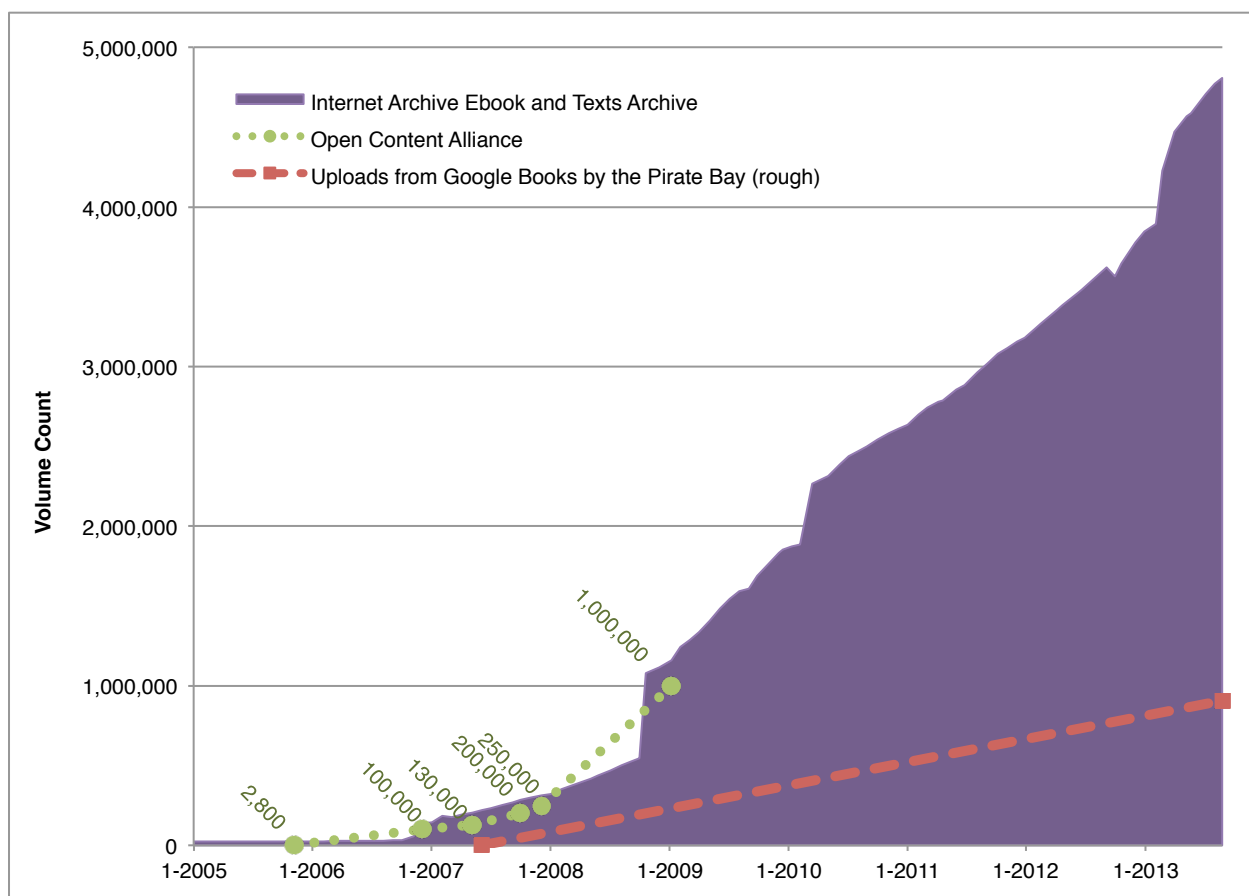


Figure 56: Growth of OCA/IA Text Collections, 2005-2013

A few other elements of this chart are also worth noting. First, while it is true that the Internet Archive was scanning and uploading books prior to the initiation of the OCA, the pace of this early scanning was miniscule compared to the periods after that. As of January 25, 2005, the IA had only posted 23,083 texts, and it would not reach 30,000 for more than a year after that. Indeed, the real acceleration in the pace of scanning seems to come only around the end of 2006, around the time the \$1 million grant from the Sloan Foundation was announced (not that there is any necessary relationship between those two events). Beyond that point, the rate of additions increases to roughly 12-15,000 volumes per month for about a year, and then increases still further, to about 20-30,000 volumes per month for the period from January to October of 2008. At the end of that period, a spike appears, reflecting the infusion of hundreds of thousands of books from the recently cancelled Microsoft scanning program. Notably, although reporting on this donation places its size at 300,000 volumes (O'Leary 2009), the observed jump in IA volume count from October to November 2008 is actually more than 530,000 – nearly twice as many. And of course, the OCA as an entity actually fizzles out somewhere not far beyond that spike, in late 2009 or early 2010. As such, while the figures beyond that point may – and very likely do – reflect

scanning by many of the same institutions that formed the OCA, that scanning was no longer operating under the umbrella of that consortial arrangement.

Taken as a whole, the IA's progress in making digital texts available, both during and after the period of the OCA, is extremely impressive, even factoring in the significant number of scans that have trickled into the Archive's collection from Google's project and elsewhere. Indeed, in a world where Google Books did not exist, the IA's five million volume collection would be by far the largest single aggregation of digital books ever created, far surpassing earlier projects like Project Gutenberg or Million Book²²⁹ (Project Gutenberg 2008, St. Clair

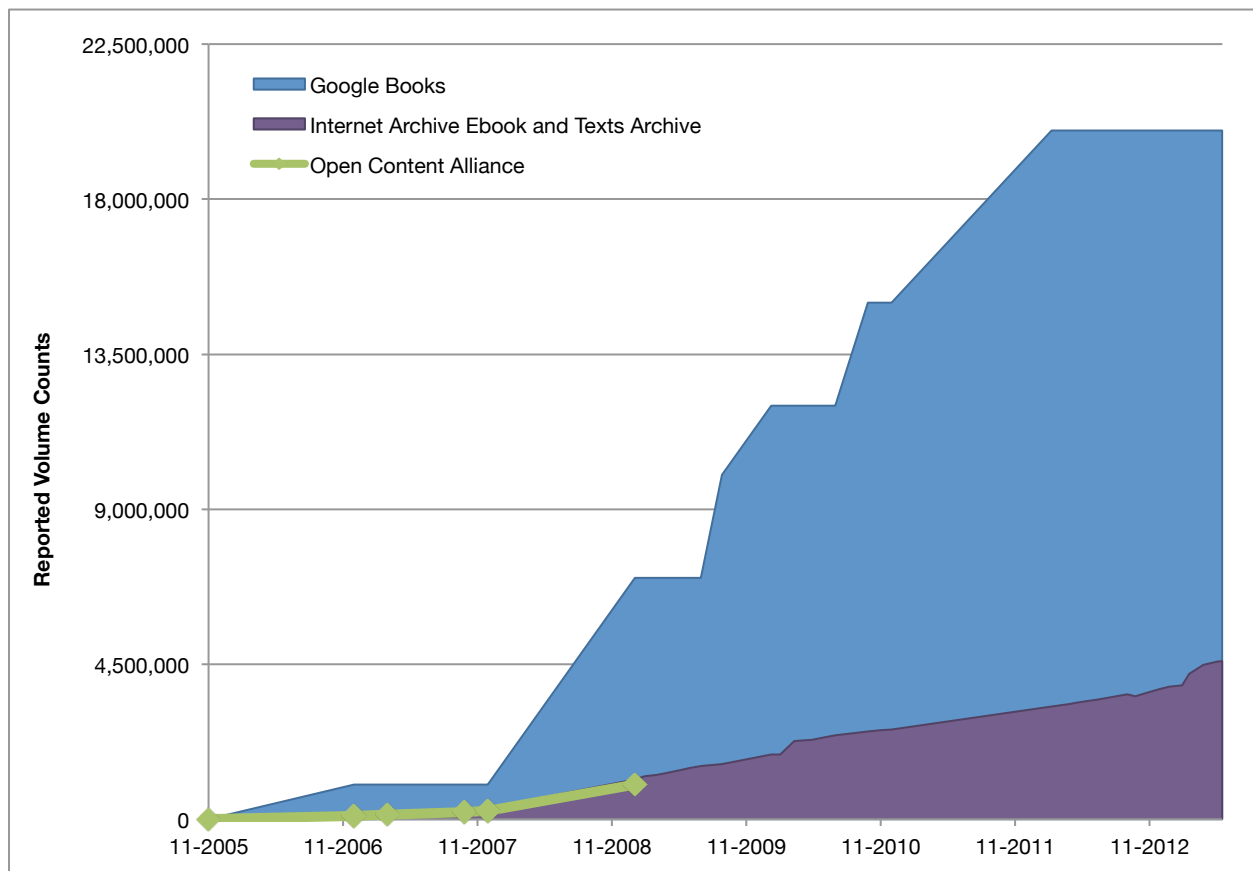


Figure 57: Growth of OCA/IA Text Collections compared to Google Books

2008, Project Gutenberg 2013, Internet Archive n.d.-b). However, of course, Google Books *does* exist. And throughout its duration, it has consistently dwarfed the Internet Archive's efforts in terms of both volume count and pace of scanning, as shown in Figure 57.

3.2.2 "Comprehensiveness"

Probing beyond both the raw numbers above and the frequent invocation of the phrase

²²⁹ Which it accomplished, in part, by absorbing these and other earlier collections whole – both of these are listed on the IA's Ebooks and Texts page as "Sub-collections," with Million Book labeled by its alternate title, "Universal Library" (Internet Archive n.d.-b).

“Universal Access to All Knowledge” in both public statements and interview accounts surrounding the OCA and IA (P1, P5, P7, P11, Benson 2005, Hogge 2005, Kahle 2005a, Kahle 2007b, Internet Archive), a more philosophical question emerges, as it did in the Google Books case: that is, how did participants in the OCA actually conceive of what it would mean to amass a “comprehensive” or even “universal” collection? Did they truly believe that “universal access to all knowledge” – or to all books – was possible, or even desirable? Having posed this question to all of the OCA interviewees just as I did with those involved in Google Books, a similar picture emerges across both cases. That is, within the OCA as within Google Books, comprehensiveness seems to have served more as an idealistic aspiration than a realistic goal. Still, there are some interesting differences in nuance within this broadly similar picture, worthy of discussion, which can be divided into three basic pieces. First, unlike the Google Books interviewees, several OCA participants expressed a high level of confidence that the creation of a universal digital book collection *was* in fact practically feasible, and not merely an optimistic ideal – that it was only a matter of mustering the collective will to do it and then keeping at it for a long time. Still – and secondly – despite this optimism about the possibility of comprehensiveness in general, a few participants were quite vehement that neither the OCA nor the Internet Archive would ever be able to achieve it. And finally, several participants made a quick pivot from the question of “all” to the question of “enough,” generally citing some numerical threshold beyond which they would be satisfied with the project’s progress.²³⁰

3.2.2.1 *Yes, Universality is Possible...*

When asked directly about the practical possibility of building a comprehensive digital book collection in general, four participants – two from the Internet Archive and two librarians – responded decisively in the affirmative, though each offered their own unique view of what that possibility meant to them (P7, P11, P14, Kahle 2011). Given the diversity within these responses, it will be useful to directly include portions of each (with apologies for the succession of lengthy quotations). At the most optimistic end, one librarian expressed a great deal of faith in the possibility of universal digitization, but notes several philosophical and pragmatic challenges such an effort would face, as well as the length of time it would take:

Well everything's possible! [...] So, you could certainly identify all the books out there and conceivably get them all digitized. But, you know, is that all knowledge? You start to get philosophical about that. So at some point you have to say, you know, this is our best representation of knowledge as we know it. I think, you know, the idea of having all ideas

²³⁰ Two participants also replied to the question of the possibility of universality with a relatively flat “no” – but as one of these was already cited in this piece of the Google Books chapter, having been involved with both projects (P2), and the other quickly elaborated her “no” into a discussion of what might be “enough” (P15), it did not seem worth including a separate discussion of this perspective here.

accessible is an interesting idea. But how do you quantify that? How do you qualify it? It's certainly, absolutely, within our means to get the public domain digitized and online. It's certainly within our means to identify, ah, you know to make our way through the orphan works. And get materials that are... that we can scan, digitized, and online. It's certainly within our means to start making agreements with in-copyright material copyright holders, and getting those materials online in some way, shape, or form. Sure! You know, it's all doable. It's time though (P11).

Similarly optimistic, though with a more quantitative orientation to defining “everything,” Brewster Kahle responded by rhapsodizing about the low current cost of digitization relative to other library budgetary demands, as well as the progress that had already been made within the Internet Archive:

Yeah! Yeah. It's cheap. So, where we are, if we take just books... um... if you take a Boston Public Library or Yale or Princeton... they're about a 10 million volume collection. That's kind of a world class library... is kind of in that... around that. Um... about 20% of it's public domain, bout 70% of it's out of print, and about 10% of it's in print. Um... to digitize a book costs currently about \$30. So a 10 million book collection would cost \$300 million. This is... oh how long... six months or so of budget of the Library of Congress. It's... we spend \$12 billion a year on our library system in the United States. 3 to 4 billion dollars a year gets spent on acquisitions. So \$300 million's tiny. Tiny. And, we've made real progress. Uh... we've got... I'd say we basically have the public domain now. There's more---cleanup to go. But, on the Internet Archive site there's about 3 million volumes. And, so I... ya know, there's some things to do, there's always cleanup to go, but the public domain, I'd say at this point, check. Now it's really working on the out of print and in print. [...] So, is it practical? Sure (2011).

Building on both of these depictions, a second librarian drew an analogy to earlier library automation projects, which were generally perceived by libraries as insurmountable tasks – but only until they were actually begun:

Yes. In terms of digital books, I think yes. [Brewster's] dream is so vast that, um... it, you know, it's kind of mind-blowing. On the other hand, that's always been the dream of people who gathered libraries, and he's done more to accomplish it on the Internet than anybody, really. At a fairly modest cost. The published world is not a very big universe. If you look at all the other things, all the other kinds of information there is. I'm reminded when I first came to University of Toronto, I had worked at Columbia, and we at that point did not believe that catalogs could be retrospectively converted. Because it was too big a job. To convert everything into MARC format. And, we actually I think at that point had... did we have a Dewey system AND a Library of Congress System? But, we weren't ever going to touch the old system. So the dream was actually so small. But I think that was what everybody felt in the United States, and that's what everyone said: it was just too costly, we couldn't do it. I came to Toronto, and they didn't know that it was too costly. And so they had actually already done it! They just set about doing it. And, um, that's why we were kind of a leader in the online catalog. Our catalog was closed in 1976, which is Library of Con-... because... we had heard that the Library of Congress was going to close it in 1976, but then they thought it was too ambitious a project. So they didn't do that. But because we thought they were going to do it, we thought we had

to do ours to be up to speed. So we did it! [laughs]. So, um, it's kind of like a tiny parallel to what Brewster thinks lots of things can be done that a lot of other people don't think can be done. But I'm convinced that, um, what has been published in books that is still extant I think is not a huge job (P14).

And finally, another IA staffer took a more philosophical tone, building upon the duration issue noted by the first librarian, and adding a more cautious note on the issue of cost coverage:

theoretically, yes. Um. Then you just need to work out how long it's going to take. But theoretically, it would be possible to put one of our scanners in every library in the world. And every publishing house in the world. And make sure that we had a copy of everything that was ever published. I mean... you know... but, it sort of becomes a fiscal issue at some point, you know? Brewster won't be able to pay for this forever, presumably, um, so, like every nonprofit struggle for a budget to keep stuff like this going... um... I guess it comes down to that in the end. Is it... financially sustainable. You know? Uh... I don't know (P7).

The high level of enthusiasm in all of these passages is notable. To a great extent, these accounts give the impression that the only thing standing in the way of universal digitization is a lack of will: given enough time, money, and faith in the possibility, these participants clearly believe that universal digitization can (and will) be achieved, and might even be attainable in no small part through the IA's existing strategies.

3.2.2.2 ...But Not Through This Project Alone

Still, a few participants were adamant in asserting that even if comprehensiveness were possible, they did not believe that this particular project would achieve it – or indeed, that it ever could have done so. As one librarian tersely opined, “*I don't see OCA having the resources or the structure. [...] You know, sadly I think they missed their opportunity. I think, uh, early. They messed up early*” (P3). And another expressed a similar pessimism in softer terms, replying,

Well it's too small in scale, I think. I think that the hope for them is handling material that maybe Google is sort of bypassing right now. Or institutions that Google is bypassing right now. Or won't work with Google. Or those that feel uncomfortable dealing with a for-profit company, and would rather deal with a non-profit. Or riskier things that a non-profit can get away with because who's going to sue them (P18).

Ultimately, he concludes, the IA efforts were “*too small in scale to get anything in a comprehensive way,*” explaining:

This stuff is really, really expensive. And Google's doing it at a scale that no one else is. We were bound by non-disclosure, but since we were dealing with both parties, we had a chance to observe both, on how they were doing it. And Google, just by an earlier start and a bigger start and more resources thrown at things and looking at it from a bigger way was evolving much quicker, than Internet Archive was in the way they were doing book scanning. Their productivity was much higher. Bigger-scale, digitizing more books,

in a shorter period of time per scanning unit, means that they're going to have a lot more say on what's going on (P18).

Particularly operating in the shadow of Google Books, the OCA appeared too small-scale, too boutique, and too under-financed to ever come close to even the still-incomplete level of comprehensiveness that Google's parallel initiative has achieved.

3.2.2.3 *Getting "Enough"*

The final facet of comprehensiveness expressed among OCA participants – the pivot from “all” to “enough” – is perhaps the most revealing with regard to the project as a whole. This sort of move is most clearly evident in one IA staffer's response to the question about whether it was possible to get “all the books:”

Yes, um...and no, at the same time. I think “all” is relative. [...] If we have a million books up online and there's estimates that it's as high as 100, 125 million that are published, is that enough? For most people a million is more than enough. They'll never go through them. Did we get the last one? Did we actually get “all”? No. Um, but we are complete enough to effective? Yes. So that to me is the exciting part of as we drive to go forward. [...] But we think after we get a certain critical mass, tipping point, um, we have “all.” We'll always keep adding to “all” but we have “all” (P5).

This, of course, is very similar to the concept of critical mass introduced in the parallel discussion of Google's relative comprehensiveness in the previous chapter – a state nearly indistinguishable from comprehensive coverage at the user level. And it is directly echoed in other OCA interviewee statements as well (P7, Kahle 2011).

However, in the OCA, this concept of “critical mass,” defined from a user's perspective, seems not to have been the only definition of “enough” at play. Rather, there is also a sense in several statements, especially by Kahle himself and by others at the Internet Archive, that “enough” might indicate some actual, preset numerical threshold; less a matter of sufficiency from a user's perspective, and more a matter of having gathered together a particular-sized pile of “stuff.” In fact, when I asked Kahle how he thought he would know when he had “everything,” he immediately responded, “We're shooting for 10 million.” But 10 million is a far smaller number than, say, Google's estimate of how many books there are in the world – in fact, it is less than ten percent of that number (Taycher 2010). And in fact, it is only half the number that Google had already announced it had scanned at the time of the interview. So one wonders what the magic in 10 million volumes might be, from a user perspective, aside from being a nice, round number.

3.2.3 Composition

There is a significant extent to which “the composition of the OCA collection” is simply a nonsense phrase. After all, the OCA no longer exists, if it ever did, and even when it was still underway, there was never any central repository dedicated to collecting scans from

that project *alone*, such that its particular shape and contribution could be assessed separately from other scanning projects. The processes and power dynamics that surrounded the selection of items for OCA scanning will be discussed in the next section; at this point, it will be sufficient to note the two major differences in inclusion/exclusion criteria between the OCA and Google Books – namely, format and copyright – as well as, more broadly, the extent to which neither the OCA nor the IA has ever had any unified strategic direction with regard to collection composition.²³¹

3.2.3.1 *Format*

The question of what types of materials would be accepted for scanning was a key area in which the Internet Archive proudly and loudly differentiated the scanning process it brought to the Open Content Alliance from that offered by Google. Many types of materials that were excluded by Google based on format – serials, fold-outs, fragile materials – as well as others never considered by Google – especially microfilm and audiovisual materials – were cheerfully welcomed for digitization by the IA (Consortium of Academic and Research Libraries in Illinois 2007, Kahle 2007b, Library of Congress 2008, Miller 2012). The emphasis on the codex format was thus much diminished (as seems fitting, given that the IA had been digitizing many other formats before it started in on books). Although clearly some of the same concerns emerged with the Scribe scanning system as with the Google scanning machines – especially with regard to the sheer physical size of the materials vs. that of the scanning bed – the Internet Archive was much more willing to find ways of working around these obstacles than Google appears to have been (Miller 2012).

In part, this greater inclusiveness relates to the different way in which duplication comes up in the OCA – that is, the libraries involved trying not to duplicate *what Google is doing*, as opposed to Google’s own more internally-focused attempt to avoid redundant scanning within its own project (an effort never apparently made by either OCA or IA). This comes up in a few places in the primary source documentation. For example, a 2007 budget request from CARLI (an Illinois library consortium) states flat out that, “Given other mass digitization projects underway elsewhere (e.g., Google’s), we want to focus our OCA content on unique material and material for which OCA’s principles of open access are particularly urgent” (Consortium of Academic and Research Libraries in Illinois). And a Boston librarian echoes this point in an interview, expressing a wish for greater knowledge about what Google and IA had already done, so that the Boston Library Consortium’s resources could be used more effectively:

²³¹ Indeed, there is a significant extent to which one might question the status of the IA’s Ebook and Texts Archive as a “collection” at all, at least in the more formal sense of the word common in library contexts: lacking any cohesive sense of purpose or direction, and allowing contributions of anything from anywhere and anyone, it is more a pile of almost entirely random donations, which in aggregate may or may not rise to the moniker “collection,” depending on how strictly the term is defined.

Between the Google content that's been done and the Archive stuff that's been done, you know honestly what we need is a master pick list of... of all the books that we target. And then some assessment of what's been done already. There's a huge amount of duplication in the Archive. We went through this in Boston. Even with the twenty libraries that were digitizing together, it was impossible for us to mash our records together and see who had what and call things. And the Biodiversity Heritage Library, that had a lot more money and had been working on that problem of de-duplication and digitizing for a lot longer, they still don't have a good way to do it. So it's very difficult to... to do those things. And so instead I'm sure we digitized some books ten times. And I know that, because when I went to pull a ceremonial book to be the first book that we would digitize in Boston, and the first three books that I pulled, I looked up in Google and they'd all been digitized already. So. De-duplication, it's a big problem (P11).

Lacking this sort of fine-grained data on what had already been done, the decision by OCA participants to venture into formats that they knew Google would not touch provided a clear way to orient their scanning efforts in a more efficient, non-redundant direction.

3.2.3.2 Copyright

The other major point of differentiation between collection policies between the OCA and Google had to do with copyright. That is, where Google had been willing to scan items in copyright if libraries were willing to contribute those items, the OCA explicitly limited itself to items out of copyright or for which copyright holders had granted permission for scanning (whether directly or through mechanisms such as Creative Commons licensing). As Kahle wrote in his initial announcement of the project,

In order to respect the rights of copyright holders, content under copyright will be made available through the OCA only with the copyright holders' authorization. At the option of the copyright holder, copyrighted content may be distributed through a Creative Commons license (Kahle 2005a).

Given the current length of copyright terms, combined with the explosion in the scale of book publishing in the twentieth century, this policy had the immediate effect of drastically limiting the “comprehensiveness” of any collection the OCA could ever produce – indeed, barring a wholesale grant of permissions from all the major publishing houses of the twentieth century,²³² it set aside the majority of books now found on library shelves (see, for example, the discussion of copyright and publication dates in Lavoie, et al. 2005). And indeed, several library partners noted exactly this limitation in their interviews (P3, P14, P15). Still, this stance proved very useful as a rhetorical device, allowing the OCA to position itself as the more virtuous, less “thieving” digitization initiative in contrast to Google Books. Indeed, as noted, it even helped win the initiative the endorsement of several

²³² It should be noted that this is not necessarily impossible; indeed, Google has essentially done that with the Google Books Partner Program, and the Internet Archive has been working on licensing for in-copyright e-book lending alongside its digitization projects. However, both of these forms of negotiated access have thus far remained far from fully “open,” at least on the terms that Kahle often advocates.

publisher organizations, including the Association of Learned and Professional Society Publishers, the Association of American Publishers, and the American Association of University Presses (Carlson and Young 2005, Suber 2005).

However, it should be noted that as the OCA progressed, the IA and some of its partners began to move quite radically away from this conservative stance on copyright, via a series of collaborative innovations in digital library lending. First, in 2007, Kahle announced that the IA would be working with the Boston Public Library, the Marine Biological Laboratory at Woods Hole, and Universidad Francisco Marroquín of Guatemala to “scan out-of-print [but in-copyright] books and offer these to users through the interlibrary loan system” (Albanese 2007c, Kahle 2007b). Then, in early 2010, IA announced its BookServer project, geared toward allowing users “to buy or check out digital copies of in-copyright but out-of-print materials, the largest portion of most libraries' scanned collections,” and dovetailing with the IA's Open Library project (Hadro 2010a). And a few months later, the first 200 or so ebook versions of out-of-print, in-copyright books went live for lending via Open Library (Rapp 2010). The Open Library's in-copyright lending program has continued to expand: it began to provide ebook lending services to patrons of 150 public and academic libraries in April 2011, and garnered the unanimous endorsement of the state librarians of all 50 U.S. states in October of that same year (Kelley 2011, Rapp 2011a). Indeed, by the time of my conversation with Kahle in September of 2011, he proudly affirmed that the Internet Archive was scanning in-copyright materials “*en masse*,” and that this had been the case for “*a couple of years*” already (Kahle 2011). As of this writing, a banner at the top of the Open Library homepage boasted that the site had over 200,000 eBooks available for lending – and that number seems likely to grow.

Intriguingly, despite the clear similarities between the underlying activities required to pursue these lending efforts and the activities for which Google Books was rapidly sued – that is, the wholesale scanning of in-copyright materials, often without the permission of the copyright holder – few have seemed especially perturbed by the IA's activities in this area. Given available information, it is difficult to know why that might be. Kahle has admitted that he has not been able to obtain permission from publishers for every book, but rather seems to have proceeded on the theory that even lacking permission, he can still scan books and make them available for lending to one borrower at a time (Meadows 2013).

Perhaps the IA is selecting its books judiciously to avoid lawsuits (for example, by avoiding books published by major extant publishing houses). Or perhaps potential litigants are for some reason accepting the IA's claim of fair use where they would not accept Google's (as discussed, for example, in Grimmelman 2010, Hellman 2010). Or perhaps (even likely) the nonprofit Internet Archive simply makes a less appealing target for litigation than the massively wealthy Google. It would seem clear, however, that to the extent that the IA is scanning entire books without the permission of the copyright holder, they are in fact

engaging in *precisely* the activity that got Google sued; if one is illegal, it is hard to see how the other could be legal, and vice-versa. Although both of Google's lawsuits have now been either settled or decided, it remains interesting – and a bit bizarre – that this apparently quite similar activity has proceeded without any similar challenges being filed.²³³

3.2.3.3 *Anti-Curation*

The final distinguishing characteristic of the collecting strategy pursued by the OCA and the IA is a bit different from the two already discussed. That is, in the OCA, although there were some broad criteria for inclusion or exclusion, no overarching collection strategy was ever put in place. The previous chapter described how in the Google Books Library Project, a clear collection strategy – albeit a very encompassing one – emerged as the project progressed: that is, the plan was to scan the full collections of a few big libraries first, then fill in around the edges with more limited subject collections scanned from elsewhere in a targeted way. Whatever the pros or cons of Google's approach, it is at least content-oriented: the idea behind it was to approximate complete coverage of the myriad subjects covered by book collections worldwide, to the extent that such a feat could ever be accomplished – and all with a minimum of duplicated effort. Within the OCA, on the other hand, there was no such orientation toward strategizing for content-based comprehensiveness. I would suggest that this was so for two core reasons: first, the particular sort of anti-establishment attitude about information access prevalent within the Internet Archive, and second, the vastly different power dynamics among the OCA partner institutions vs. between Google and its partner libraries.

My interviews with individuals connected to the Internet Archive's various book initiatives convey a strong impression that the prevailing attitude toward librarians and traditional library practice within the IA was, if not hostile, then at least deeply dubious.²³⁴ Within these interviews, library systems were often depicted as the old, outmoded way of doing things, where the Internet Archive's systems were the future. For example, librarians were depicted as tending to obstruct the flow of knowledge, where the IA was trying to free it (especially Kahle 2011); as being unnecessarily finicky about the media they choose to collect, keep, or digitize, where the IA welcomed all possible content (P5); and as being overly wedded to top-down classification and metadata schemas, where IA would open all

²³³ To be clear, I think the Internet Archive's digital lending program is a great idea, and I hope they don't get sued. It has just been a source of continual surprise to me that the Author's Guild in particular – those folks who are still pursuing a series of appeals against not only Google, but also HathiTrust – has not latched onto this program with their usual brand of righteous indignation.

²³⁴ Some specific librarians were exempted from this attitude of suspicion: in particular, the University of Toronto's Carole Moore is cited frequently as one of the rare librarians whose actions and principles regarding information access fall closely in line with the IA's and with Kahle's (P2, "Milestone Achieved" 2006, Kahle 2011).

such schemas up to Wiki-fication (P7).²³⁵ Overall, IA staffers seemed to relish the idea of creating a new kind of library – one without *librarians*, except possibly as contributors of content in the early stages. As one of these individuals explained with pride,

it's a liberated position to work at Open Library because we're not beholden to the normal baggage that is associated with library systems and library practice, and none of us are librarians. We're software people. You know? And that allows us to sort of approach the issue in a totally fresh way. And you know I guess I'm pretty proud of that, that you know it's deliberately NOT a traditional library catalog (P7).

And admittedly, there is something appealing there, in tossing off the accumulated detritus of the library professions and starting fresh. However, with regard to collections, it soon becomes clear that one piece of “baggage” that was jettisoned – for better or worse – was the traditional library focus on collection development and curation; the longstanding practice of purposively selecting sets of material to meet a content-oriented need or ideal.

To my knowledge, neither Kahle nor anyone else at IA has ever expressed any particular interest in controlling or even influencing what the partner libraries were scanning, within the OCA or beyond it. Whatever was scanned was more information, and more information was better. Assessing the quality and value of information, ensuring breadth, avoiding duplication – those activities were part of the old way; obstacles to progress. If libraries insisted on performing them (as indeed they did (P5)), they were on their own. The new way – the IA way – would be inclusive, democratic, non-judgmental, and primarily interested in gathering together as many pieces of information as possible, regardless of the actual informational content thus assembled.

Still, the anti-establishment attitude toward libraries within the IA fits somewhat uncomfortably alongside the other core reason that no centralized collection development strategy was ever undertaken within the OCA. That is, among the OCA partnerships, no individual institution wielded sufficient authoritative or allocative resources for the other institutions involved to fall in line behind it and actually coordinate their activities. Consider the distribution of the sources of power among the OCA partners as compared to the distribution within the GBLP. In the GBLP, Google held most of the cards. It had the billions to pay for the scanning (or at least to cover the vast majority of the costs), and was willing to do so. It also had a massive public profile, with which librarians were interested in aligning their institutions in order to improve their reputations and make them seem more modern. Those are significant resources to bring to bear on a project like this. Google's library partners, by contrast, were relatively weak, especially after the first five contracts were signed – the most any library could realistically do beyond that point, should they not care for Google's terms, was to refuse to sign on. And then Google would

²³⁵ And, it should be noted, this is all despite the Archive's legal status as a library (Kahle 2007a), and Kahle's adoption of the “librarian” title for himself (a sign declaring as much hangs above his office door).

simply move on to the next library. Given this large imbalance of power in favor of Google, it is not surprising that Google was able to decide on a collection development strategy and simply follow it, regardless of the specific interests of its library partners.²³⁶ No other entities in the program wielded anywhere near enough clout to shove that trajectory off track, even had they wished to do so.

In the OCA, however, the situation was different. No central figure was providing all of the funding – far from it. Microsoft spent the most, but still only contributed a very minor amount compared to what Google was willing to spend (Guess 2008). And there was never any question of the Internet Archive picking up any large percentage of the tab – it was simply not wealthy enough to do so (e.g., P3, P7). Instead, much of the funding for the scanning that was accomplished through the OCA came through the libraries themselves – whether from internal funds or external grants – in line with Kahle’s consistent vision of how such projects should proceed (e.g., P14, Kahle 2011). And this put each individual library in a considerably different – and much more enfranchised – position with regard to setting scanning priorities than was the case within the Google partnerships. That is, where Google alone had the final say over what would be scanned through its efforts, each OCA library operated under its own discretion, and according to what its internal management or external funders were willing to bankroll. As one librarian explained, at the University of Toronto,

the order [was] basically based on who was willing to fund what. So we are interested in everything. But, um, different funders had different priorities. So it was either Microsoft wanted to do English, thinking that everything was in English (interestingly!). Some of the materials that have been used the most from our collections have been in Chinese and Ukrainian, and languages and materials that are not easily accessible around the world. And some of our Medieval materials as well. We have a really good Medieval collection. Um... And, the Medievalists... People are interested in Medieval studies everywhere. It really is fascinating. Um... so, some of those collections have been done because they're special interest groups usually. Perhaps, communities, local communities, who have immigrated here wanting to represent their heritage. And certainly we got a large grant from the, um... to do French material to support French education, in Ontario, actually. So... It's not logical from the point of view of we just work through my priorities, or... but it's... it's probably better that somebody has these priorities and they're willing to put up the money for it (P14).

Similar stories, involving varying levels of funding and a wide range of subject areas – from Yahoo’s initial contribution of \$150,000 for scanning American Literature at the University of California to a \$4 million grant for the Biodiversity Heritage Library to scan taxonomy materials – abound throughout the available data from OCA partner institutions (P1, P3,

²³⁶ Google’s unilateral decision to stop scanning duplicate items from libraries, even where it had agreed to scan entire collections (i.e. at Michigan and Stanford) is one clear example of how this power imbalance played out in the area of collection strategy. This is described in detail in the previous chapter, section 3.2.3.2.3.

P11, P15, P18, "Internet Archive Receives Grant" 2006). Yet, with scanning priorities set exclusively at the local level in this way, no OCA partner institution was at any point required to consider the objectives of the greater whole. The Internet Archive was willing to take anything, and they weren't the ones paying for it anyway. And so at the level of the collective depository, idiosyncrasy reigned.

4. Implementation

As in the previous chapters, the final section here will describe the processes and structures through which the OCA was implemented. However, it should be noted at the outset that this is a far less straightforward task for this case than for any of the other three, for one very basic reason: at the level of implementation, the Open Content Alliance as such never actually cohered. Recall, for example, that some of the individuals most involved in the leadership of the OCA now claim that the OCA never existed (P1, P11); the project simply never reached the level of a functioning, fully-implemented initiative. Thus, while for the purposes of this analysis, it is possible to cast some light on a few corners of the OCA's patchwork of implementations, the sense of an overarching pattern or design that emerged in the previous three cases does not come together as clearly here. And although that may be somewhat narratively unsatisfying, it reflects the reality of the case. Pieces of things happened – in some cases impressive and marvelous pieces – but those pieces never came together in the way they were intended to have done, under the banner of the OCA.

4.1 Processes

This section will discuss four elements of the OCA's implementation patchwork: (1) the fundamental and destructive conflict among OCA leadership over administration and governance of the project, (2) the mechanics of the actual process(es) through which books and other items were digitized by OCA partners, (3) the lack of clarity regarding end-use constraints for materials made available through IA (both digitized as part of the OCA and otherwise), and finally, (4) the divergence between rhetoric and practice with regard to transparency about the inner workings of the project.

4.1.1 Administrative Structure (or lack thereof)

Many of the project's leaders assert that in retrospect, the OCA's failure to cohere and thrive as an independent entity most likely resulted in large part from a fundamental disagreement among the leadership over whether the OCA should stand apart as a self-sufficient organization (as many librarians desired) or continue on as a subsidiary organ of the Internet Archive (as Kahle preferred). As one early participant explains, "*one of the things that became really clear was that many of [the OCA's] members – well...it wasn't even 'members' it was 'supporters' – wanted an organization;*" specifically, he continues,

Tom [Garnett, of the Smithsonian Institution] very much thought that we should have an organization. He didn't like the idea that, um, the OCA was being, was kind of a

project of the Internet Archive, and ended up staying that way. He wanted more. He wanted it to um, engage in advocacy, he wanted book scanning money to go through the OCA, Brewster didn't. Brewster wanted it to go through the Internet Archive....Um...I think Brewster wanted the OCA...name to continue to be an Internet Archive thing. But basically he saw it as a project of the Internet Archive rather than, than some entity to which the Internet Archive was simply a member (P1).

Several other participant accounts echo this assessment of the conflict between Kahle and other OCA contributors in the area of governance, noting Kahle's resistance to governance in general (P3), as well as the OCA's failure to live up to many participants' organizational expectations (P2). Ultimately, of course, Kahle got his way: the OCA remained subordinate to the Internet Archive, to such an extent that it dissolved and diffused into the IA's other activities, all but disappearing itself. As one librarian observed in late 2011,

you can't separate the OCA from the IA. First of all. Because there is no OCA. You know, and Brewster started to use the word again, and for a time he agreed to a moratorium on the phrase because he understood that he had kind of pulled a fast one with... with what happened. But, there is no OCA. There is no organization that is thriving, and that has a community spirit, and so on. There is, very much, though, a scanning program run by the exceptionally... exceptionally well by the Internet Archive. And... that helps libraries digitize their collections under open principles. So, you know, it's really good at that. It's really, really, really good at that (P11).

Looking back, a few participants do portray the lack of an additional governance layer as a liberating factor, allowing for greater nimbleness and flexibility among partner institutions (P7, P14). And yet, for the OCA, this nimbleness came at a significant cost in both leadership and funding – significant enough that it likely sounded the death knell for the initiative as a whole. That is, as noted, the subordination of the OCA to the IA was a primary reason for Maura Marx's decision to abdicate the OCA leadership role for which she had been hired, and instead to embark on two other Sloan-funded initiatives in the same space – the Open Knowledge Commons, and later the Digital Public Library of America (P1, P3, P11). In combination with Microsoft's termination of its scanning program and retraction of funding around the same time, this unfortunate resolution to the governance question dealt a crushing blow, from which the Alliance never recovered.

4.1.2 Scanning

During the period when the OCA was still functional, the actual process of moving books from stacks to screens among its contributors – like the process of aggregating its collection as described above – was not singular or linear, but plural and idiosyncratic. From the outset, each library had the freedom to develop its own particular procedure for scanning, execute that procedure, and contribute the resulting scans to a central repository or set of repositories. Many – perhaps even most – chose to contract much of the actual scanning out to the Internet Archive, but not all did (e.g., Carlson and Young 2005, Young 2006a). And

within the Internet Archive, a standardized procedure for scanning library collections did eventually emerge, but it is unclear how much of this process was actually in place during the lifespan of the OCA.²³⁷ Still, lacking more precise historical accounts of the process, some understanding can still be gained by examining the current specifications, which are likely not wholly dissimilar to earlier permutations.

As it turns out, the IA's standardized scanning process has striking similarities to Google's, with only a few significant divergences. Thus, as in previous such cases, this discussion will mainly focus on the elements that differ, while only briefly noting the similarities. As a starting point, Figure 58 depicts a simplified version of the current Internet Archive book digitization workflow, drawn from IA's publicly available documentation for potential digitization partners (Miller 2012, "Content Agreement" 2013). To the extent scanning is handled through IA, the overall flow of this process, as the left side of the figure shows, works almost exactly like Google's: books (and other materials) get loaded up and shipped off to scanning centers, where they are evaluated, scanned (or not), and sent back, and then a digital copy follows at a later point.

²³⁷ The Partner Documents page on the IA website (<http://archive.org/details/partnerdocs>) was first archived by the Wayback Machine in April of 2010, and at that time it did not include any formal book scanning process document. The current document (Miller 2012) states in the footer that it was created in June 2009, but no prior versions are available either directly on archive.org or via the Wayback Machine. As such, it is impossible to say what information it actually contained when first posted, and thus how standardized the process actually was at that point.

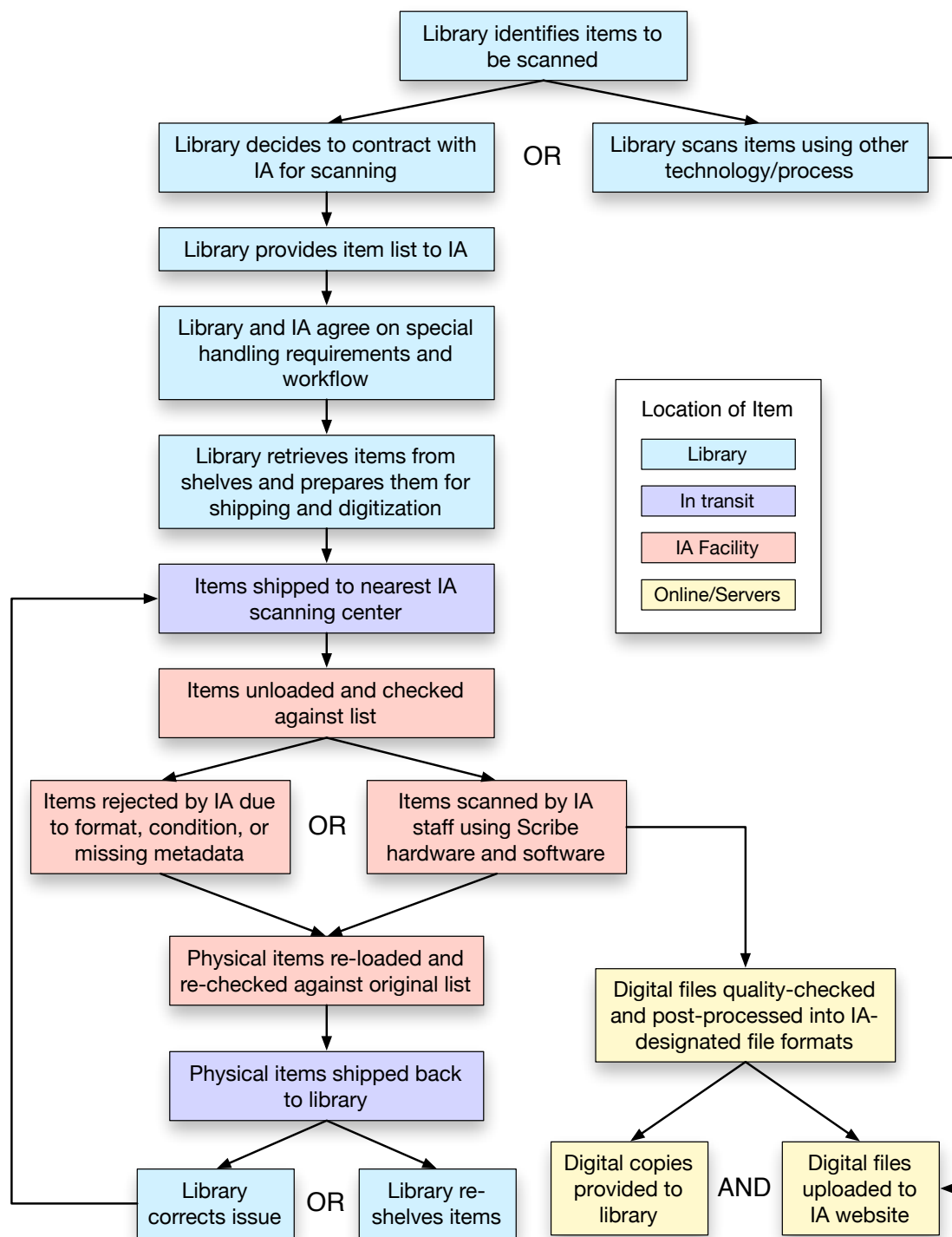


Figure 58: Internet Archive Book Digitization Workflow as of 2012.

Still, although the overall process is quite similar, several of the component parts function somewhat differently. One of these – the greater breadth of formats IA was (and is) willing to scan – has already been discussed. In addition to standard bound monographs, IA has facilities for scanning microform, folios, archival collections, photographic materials, loose-

leaf materials, and audiovisual materials (Miller 2012, 15); hence the use of the word “items” rather than “books” in the figure. Unlike Google, IA is also willing to digitize materials that require special handling, so long as that service is negotiated in advance. And further, the issue of advance negotiations raises another divergence from Google’s processes; one that underlies the workflow above, but does not directly appear within it: that is, the question of cost coverage. Where Google’s contracts specified that Google itself would cover the costs for transporting and scanning library materials, in the Internet Archive’s process these services are either paid for directly by the library to IA or invoiced by IA to the library upon completion, at a rate negotiated in advance based on the materials to be scanned (Miller 2012, “Content Agreement” 2013).²³⁸

Additionally, it is worth reemphasizing that all of the above only describes the left side of Figure 58, depicting the process of outsourcing scanning to the Internet Archive. The right side, however, illustrates the fact that the libraries involved in the OCA were never actually required to use that process at all. If a library already had its own scanning apparatus, or its own system for outsourcing scanning, or for that matter, collections already digitized and ready to go, it could bypass the Internet Archive’s scanning system entirely, and simply contribute its own scans to the common pool. This, of course, is a further departure from Google’s procedure, which relied heavily on total standardization and process control, and did not accept other sorts of scans from libraries (or, to my knowledge, anyone else, aside from its publisher partners (Google n.d.-c)). Interestingly, however, this free-form scanning policy seems to have had some unintended negative consequences for secondary initiatives – and specifically HathiTrust. Like the Google scans, many OCA scans are winding their way into the HathiTrust corpus. However, according to one interviewee with deep ties to that initiative, the diversity in OCA imaging and processing procedures has made it much, much harder for HathiTrust to integrate OCA files than GBLP files. In fact, this individual even went so far as to claim the Google scans were actually superior to the OCA’s and to IA’s, because they are uniform and predictable, and because problems could be dealt with through a single source (i.e. Google) – neither of which are true of the OCA’s collective output (P8).

4.1.2.1 *Operation of the Scribe Scanning Centers*

The actual mechanics of IA’s scanning centers also merit further discussion in their own right. During the period of the Open Content Alliance, the Internet Archive put eight scanning centers into operation in three countries, each equipped with 10 Scribe scanning machines, which were kept running 16 hours per day by staff hired by the Internet Archive

²³⁸ The IA’s current contract template lays out a few different pricing models from which libraries can choose. At a basic level, its book scanning runs about \$0.10 per page, plus a small setup fee per book. Other formats, including foldout pages in standard-sized books, cost extra (“Content Agreement” 2013).

(Young 2006a, A&G Total Media 2007). All of these scanning centers were located in or near early OCA partner institutions; for example, the Boston scanning center was (and is) located in the Boston Public Library, the Washington DC scanning center in the Library of Congress, and the Toronto scanning center in the University of Toronto Library (Young 2006a, "Boston Library Consortium and Open Content Alliance to Provide Digitized Books" 2007, Library of Congress 2008). The full list of scanning centers in operation as of 2007 appears in the table at right. Since

Early Scanning Centers
Richmond, CA
Los Angeles, CA
Urbana-Champaign, IL
Toronto, ON
Washington, DC
New York, NY
Boston, MA
London, England

that time, the list of IA scanning centers has changed a great deal, though it is unclear when each change occurred. The current IA documentation lists seven "regional Internet Archive Scanning Centers," among which, three are the same as in 2007 (Boston, MA; Washington, DC; and Toronto, ON), while four are different (San Francisco, CA; Fort Wayne, IN; Princeton, NJ; and Shenzhen, China). In addition, the current document also notes that there are 20 smaller IA digitization centers, located in "Brasilia, California, the District of Columbia, Idaho, Illinois, North Carolina, Oahu, Utah, Canada, Scotland, Guatemala and the United Kingdom" (Miller 2012, 21) – quite an expansion over the past six years.

The actual Scribe scanning machines, like Google's scanning apparatus, are non-destructive with regard to the books. However, the shape and functionality of the Scribe is a bit different from the Google machinery. As the Internet Archive's process documentation describes,

The Scribe workstation is comprised of a sturdy aluminum frame that supports two adjustable camera mounting rails, two color cameras that capture both recto and verso pages of each book, a floating V-shaped book-cradle whose angled design minimizes stress placed on materials, a glass platen that is raised and lowered by means of a foot pedal, two banks of museum grade lights that illuminate the book and one computer that captures the color images from the camera and performs some of the pre-processing. Once the book is digitized and an on-site Quality Assurance process is completed, the captured images are uploaded via RSYNC to processing computers located in California (Miller 2012, 3).

Figure 59 provides two views of the Scribe hardware in operation at the Internet Archive in San Francisco (sadly, this facility was badly damaged by fire in 2013). Essentially, where the Google system uses a flat book platform, relying primarily on software and imaging algorithms to deskew and refine the page images after the fact, the Internet Archive system relies more strongly on the hardware itself to ensure the quality of the image: the book is pressed in a v-shape to flatten the pages, which helps to reduce curvature and therefore skew, producing a cleaner image straight off the scanner, though at some cost in speed (as it takes time to raise and lower the platen glass between each page image). Still, particularly

at the time of its release in 2005, the Scribe system represented a major advance in efficient, high-quality, low-cost book scanning; although some earlier systems could match or exceed it in speed and/or quality, such scanners remained quite expensive, and thus often difficult for libraries to justify purchasing.²³⁹ Additionally, because the image is clearer directly off the scanner, there is less need for investment in post-processing of the kind that Google has made – the raw images are already very good.



Figure 59: Scribe Scanning Stations at the Internet Archive (Photos by Jason "Textfiles" Scott [CC-BY-2.0] via Wikimedia Commons)

4.1.3 Constraints on end-use

As explained in the discussion of OCA motivations above, the principles of openness espoused by the initiative – especially with regard to access and rights – were central to many libraries' desire to contribute. The perception of Google Books as restrictive in this regard, vis-à-vis both its library partners and its end users, was pervasive here (though only inconsistently grounded in fact), and the OCA was thus successfully presented as a liberating alternative. And for the libraries involved, it was unquestionably freer and more empowering: unlike Google, neither the Internet Archive nor the OCA ever placed any restrictions on what contributing institutions could do with the digital versions of items scanned from their collections – though some funders, especially Microsoft, did do so

at least for a time (P1, P15, Suber 2005).

However, for end users, the "openness" of the content scanned by the OCA and its successors at the IA has always been far more complex. And somewhat ironically, this resulted in large part from the freedom the initiative allowed its partners. That is, as previously noted, each contributor to the OCA, whether funder, scanner, or content

²³⁹ For example, I recall the Kirtas automatic scanner purchased by Northwestern University Library in 2004 (my employer at the time) costing about \$150,000; not an insignificant expense, even for a library with fairly substantial resources like Northwestern's.

provider, was allowed to set its own terms for what could be done with the content it helped place online. For example, in his announcement of the OCA, Kahle averred – in striking contradiction to his position on Google’s restrictions – that:

We believe that donors should have the option to restrict the bulk re-hosting of a substantial part of a collection. This seems fair and is similar to the Creative Commons Sampling license (Kahle 2005a).

And further, as also noted, the OCA and IA have engaged in a great deal of scanning of in-copyright materials, some with the permission of the copyright holders (for example, many institutions have scanned their school newspaper backfiles or yearbooks, to which the schools themselves generally own the rights), and some not, and it is not at all clear what rights a secondary user might have to make further use of those materials even where they are made openly accessible through IA. After all, the permission for scanning and hosting was given to *the Internet Archive*, not to the IA’s end users. Were a random individual to download these files in bulk and rehost them on another site without seeking her own permissions from each and every copyright-holder involved, it is far from clear that doing so would not expose her to significant risk of litigation. In practice, the diversity in access policies and rights statuses for materials made available through the IA’s Ebook and Texts Archive (both the OCA scans and the others completed since) created a multi-layered pastiche of access restrictions, leading to a very confusing and potentially risky situation for end-users wishing to reuse the content thus provided.

In 2008, Kalev Leetaru executed one of the few scholarly comparisons of Google Books and the OCA that has ever been done, focusing especially on issues of transparency and rights within both projects. His analysis concluded, in essence, that Google Books was more transparent and more open than it was generally thought to be, while the OCA was less transparent and less open than it was generally thought to be. And while some of the transparency issues he notes within the OCA have been remediated by IA in the years since the article was published (on which more in a moment), his assessment of the rights issues remains solidly on point. As he writes, the OCA’s policy of allowing “each of its member organizations set their own distinct rights policies, rather than enforcing a single global set of rights” effectively

forces users to inspect the rights statement of every single work they find to determine their rights to the content (such as whether it can be redistributed). Further complicating matters, the OCA entry for most works does not provide information on the rights restrictions for a given work, only the copyright status of the original printed material. To determine the rights status of any particular work, a user must follow the link in both the Digitizing Sponsor and Book Contributor metadata fields to view the rights restrictions enforced by the two organizations (Leetaru 2008).

In Google Books, the rights status of each work is made extremely clear, not only by the

uniform rights statements attached to each item, but also, in most cases, by the amount of the work made available for viewing online. By contrast, as the passage above illustrates, on the IA site each item superficially *looks* equally open (all tend to be available for full text download in a number of formats, and the rhetoric of openness pervades the site), but many may in fact be subject to a number of restrictions, copyright among them. Indeed, the IA's FAQ on Rights explicitly states that

we cannot give ironclad guarantees as to the copyright status of items in our Collections and cannot guarantee information posted on [item] details or collection pages regarding copyright or other intellectual property rights. Our terms of use...require that users make use of Internet Archive's Collections *at their own risk and ensure that such use is non-infringing and in accordance with all applicable laws* (Internet Archive n.d.-c, emphasis added).

The obligation to suss out the rights status of each individual item seems a significant burden to place on users who have come looking for "open" content. And in fact, this requirement would be particularly onerous for Kahle's favorite type of potential user – the type who want to use these materials *en masse* to build a multiplicity of new interfaces, to guard against the hegemony of a single system (e.g., Kahle 2011, Vaidhyanathan and Kahle 2011). If the barrier to such usage is ensuring the proper rights and restrictions statuses are present for each of a few million items (beyond having the technical skills to engage in interface design and data mining in the first place), that becomes a very high barrier indeed.

In sum, though at the level of principles and ideals, the OCA was strongly in favor of openness and a lack of limitations, in practice those principles seem to have extended only as far as the contributing institutions themselves. The Biodiversity Heritage Library or Microsoft or any other partner was perfectly free to do as it liked with the materials the institution itself had contributed to scanning, and in this way the OCA was certainly more open than Google Books. But because that institutional freedom extended to the freedom to restrict what others could do with those same materials, the rights situation for end-users in the OCA turned out to be considerably less comprehensible and user-friendly than the parallel policies for Google Books: regardless of whether OCA members' restrictions were less stringent than Google's or not, they required a great deal more user effort to even discover – and that in itself is a barrier to effective access.

4.1.4 Transparency in Practice

There was a disconnect between principle and implementation on the issue of transparency within the OCA, as on the issue of openness just discussed. For transparency, however, the issue was less one of confusion and complexity than it was a matter of scale and resources.

Transparency exists on many levels: it is one thing to be willing to answer questions if asked; it is quite another to actively post and publicize documentation about what you are doing in easily accessible venues. Over the years, many of the OCA's participants have

freely discussed the inner workings of the project with anyone who asked, including myself, and those person-to-person contacts are certainly one way to be transparent. But benefiting from that type of transparency requires having personal access to one of those individuals – and that is not a low bar to surmount for this project, which has provided little public information about who exactly those individuals might be. And for anyone lacking such interpersonal contact, it has always been strikingly difficult to find out much about the OCA at all. The information on the OCA’s website has always been sparse, and the information about the OCA and its successor initiatives on the IA website remains sparse as of this writing; beyond the IA’s Partner Documents page, which was only added toward the end of the OCA’s lifespan, very few details about the actual mechanics of the IA’s digitization processes have ever been provided. Leetaru, in fact, writes with some frustration on this point, noting that while the OCA “described its lack of published technical documentation as simply a matter of limited resources, a result of a small technical staff who cannot afford the time to generate detailed technical documentation for publication,” it simultaneously circulated “extremely detailed documentation, including workflows, exacting hardware and software specifications, and other key documentation...among its engineering staff.” And he goes on to conclude,

It remains unclear why OCA has not published substantive information regarding its operations, even while Google engineers offer considerable detail in the field literature. However, the reasoning behind why both initiatives chose not to publish exhaustive technical information is not truly relevant. Whether it is selective dissemination or a simple lack of time, the end result is the same: precious little technical information is available to those interested in pursuing their own large-scale book digitization campaigns (Leetaru 2008).

Leetaru also takes issue with the lack of transparency surrounding the OCA’s contracts, none of which he was able to locate at the time of his writing. Since then, a few of these contracts have been made available, from the University of California and the Boston Public Library (Internet Archive and University of California 2006, Internet Archive and Boston Public Library 2007), along with more documentation of the IA scanning process. And yet, given the rhetoric of transparency surrounding the project, it would seem more ideologically consistent for many more such documents to have been made accessible for public view, ideally posted online in easy-to-find locations, from the very beginning of the project. Why this has not been the case remains unclear.²⁴⁰

²⁴⁰ Of course, it is possible that many of the OCA’s scanning partnerships were left undefined by contracts. Although the Partner Documents page does now include a template for such agreements, it is not at all clear that this template or anything like it was in use before about 2009. In fact, a few interviewees for this case specifically noted Kahle’s distaste for contracts, and his reluctance to employ them on behalf of the IA (P3, P7); one even went so far as to state that “We [IA] try not to sign contracts with anybody. ... We refuse, as deeply as possible” (P7).

4.2 Structures

As noted in the previous chapter, digital information structures tend to be far less fixed and unitary than physical buildings like libraries. And in the case of the OCA, a certain level of multiplicity was always intended, by design; it was not meant to be “*the universal library*” so much as an effort to provide the source material for a dozen or a hundred or a thousand libraries, all with their own purposes, scales, and audiences. Still, as in other parts of this chapter, there are a few thematic threads that can be plucked from among the OCA’s diversity of infrastructural arrangements. Three of these will be discussed below: (1) the design and affordances of the two most comprehensive interfaces to the scans produced by the OCA and its successors, at Archive.org and OpenLibrary.org; (2) the extent to which OCA’s ideal of openness led to an emphasis on platforms over portals; and (3) some brief notes on the role of imaging and metadata standards in this project, especially as compared to Google Books.

4.2.1 Primary Interfaces: Archive.org and OpenLibrary.org

Although in theory anyone may be able to build an interface around the OCA book scans, in practice there are still only two sites that provide roughly comprehensive portals to all of those materials together: the Internet Archive’s Ebook and Texts Archive (archive.org/details/texts) and Open Library (openlibrary.org) – also a project of the Internet Archive. The latter is a relatively recent addition: though first envisioned in 2006 and architected in 2007 (Swartz 2007, Kniffel 2008), it only really came into its own as a full-featured site in about 2010, after the Open Content Alliance as such was already essentially dead (Hadro 2010a, Hadro 2010b, Kahle 2010d, Rapp 2010). And although Microsoft did at one point operate a third interface to these works – Live Search Books – it only did so for about 18 months, from December 2006 to May 2008 (Guren 2006, Albanese 2008d, Nadella 2008). Given the late entry of Open Library and the early exit of Live Search Books, the only interface to consistently provide access to these scans in a central location for the entire lifespan of the OCA (and beyond) has been the Internet Archive’s – and it is thus this interface that I will describe first.

The Internet Archive Ebook and Texts Archive (Figure 60) shares a design with Internet Archive site as a whole – and it is not one especially well suited to book discovery, reading, or research. The site is full of distracting animated gifs (on the live site, the book image at the upper left in Figure 60 blinks continuously, as do all book images on the site), the search feature is not full-text capable, and the overall design is fairly antiquated – and beyond all this, the underlying server infrastructure is unstable and prone to frequent bouts of downtime (P1, P3, P7). On the books site specifically, moreover, the major navigational divisions on the front page are, somewhat oddly, based not on any categorization generally thought helpful for browsing books (e.g. subject areas, fiction/nonfiction), but rather on the institutions that provided the items to the Archive (e.g. American Libraries, Canadian

Libraries, Universal Library). While there is an extent to which this could be seen as a logical extension of the “archive” aspect of IA (as archives do generally organize collections by provenance, rather than accessibility), as well as the IA’s emphasis on empowering partner institutions within its scanning projects, it does not do a great deal to aid users in navigating the resources offered on the site. There is also a list of announcements and recent additions near the top of the page, and a list of “Staff Picks” along the right, but it is not clear how frequently these are updated, or on what basis the “picks” are selected.

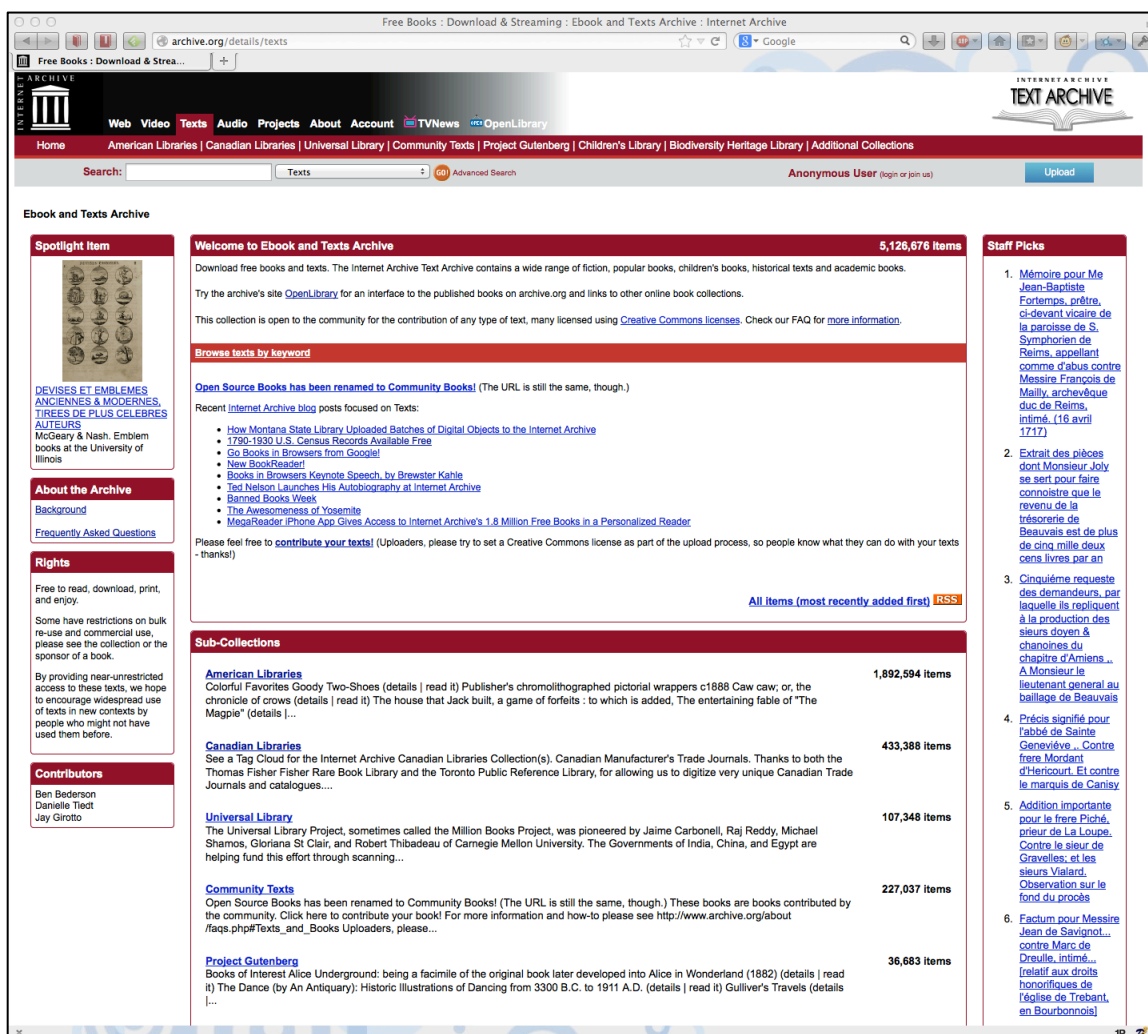


Figure 60: The front page of the Internet Archive Ebook and Texts Archive

The item record pages for books on IA (as in Figure 61) maintain much of the look and feel of the main Ebook and Texts Archive page – still a blinking book in the upper left, and similar navigation. The full text of each book²⁴¹ is made available for download in several

²⁴¹ On the IA Ebook and Text Archive site, the books are all full text; this differs from the Open Library site, which includes many records for books which are not available on Open Library in full text. The IA books represent a full-text subset of the Open Library records.

different formats – PDF, EPUB, Kindle, Daisy, plain text, and DjVu – either individually or all together as a torrent. These options appear just below the image of the book. Beside those links, the metadata near the top resembles that found in a library catalog or on Google Books, though it does explicitly note the book’s “Possible Copyright Status” (in this case, “NOT_IN_COPYRIGHT”) as well. Toward the bottom, a set of starred reviews occupy much of the page space – a feature that one IA collaborator pointed to as problematic, as it seemed inappropriately judgmental, and more suited to an entertainment site than one geared toward research and discovery (P1).

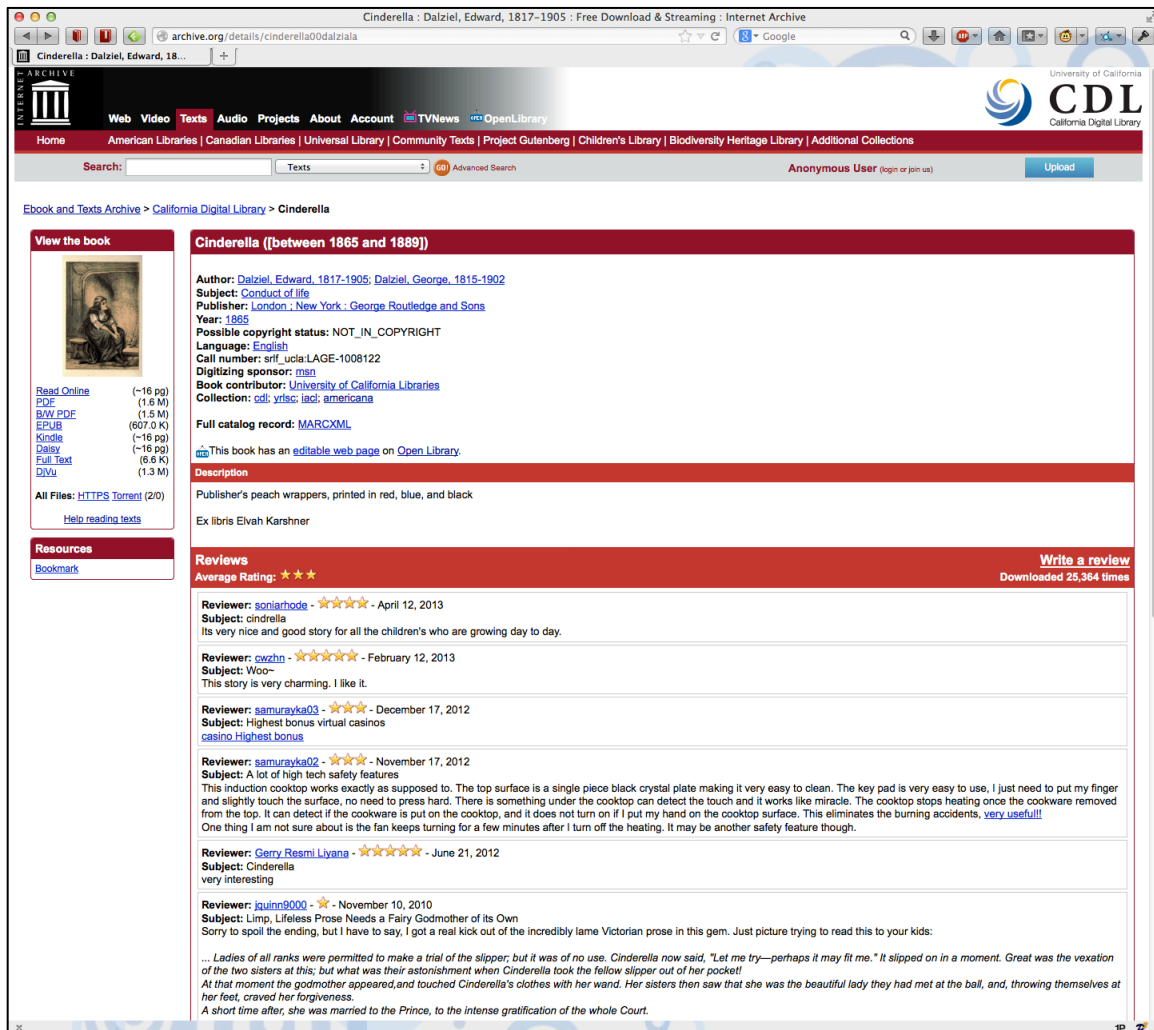


Figure 61: Item record for an edition of *Cinderella* on the Internet Archive EBook and Texts Archive (scanned by Microsoft from the University of California)

The IA item record also provides an option to read the book online, and the interface it provides for doing so creates a bridge between that site and the Open Library site, as the same digital reading interface is used by both. The Open Library reader, shown in Figure 62, was the first piece of the Open Library site to go live as a demo, and already had much the same look and feel it has now as early as 2007 (Kahle 2007b). It provides a quite

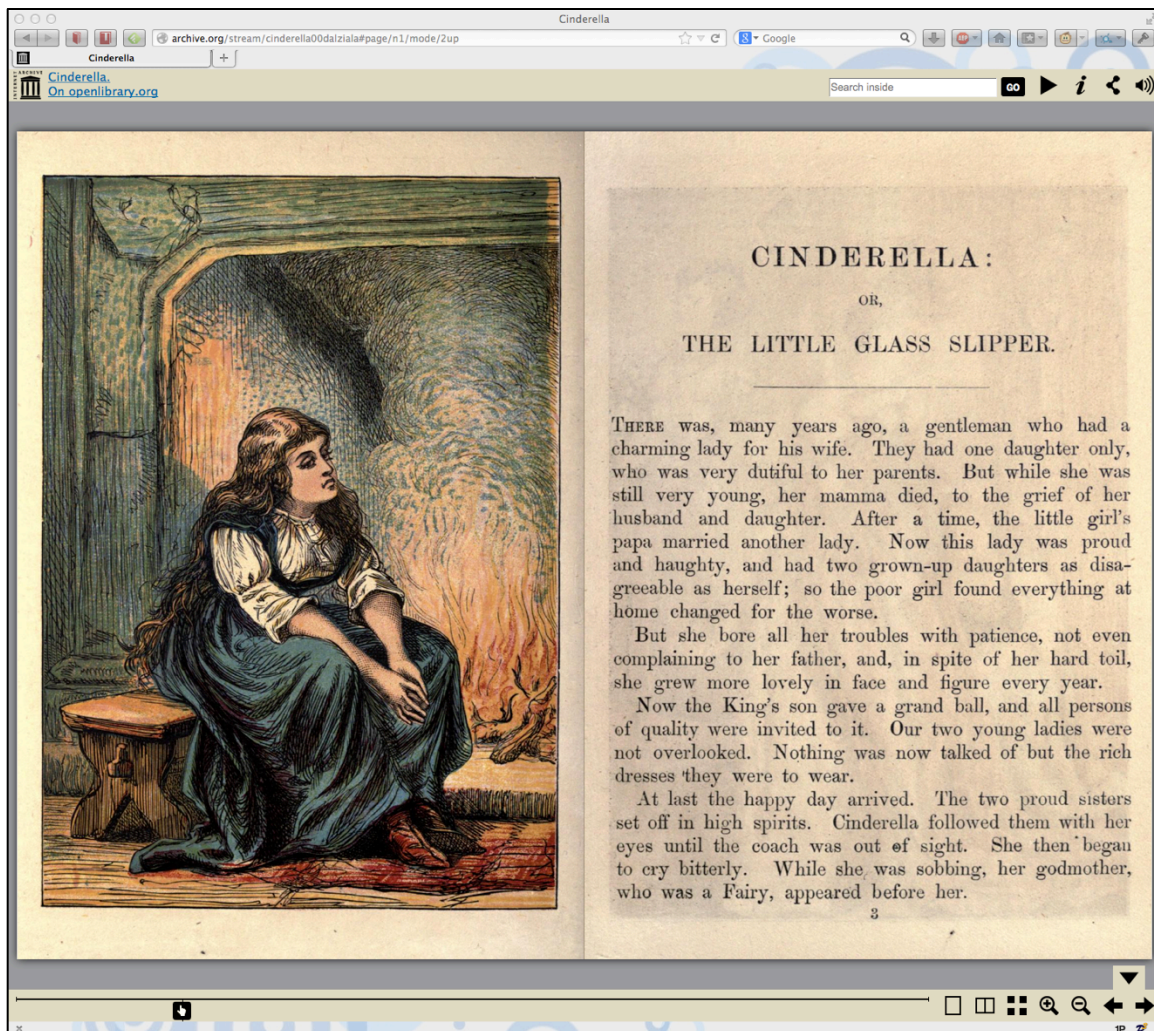


Figure 62: *Cinderella* in the Open Library Book Reader

pleasant, very book-like user experience, complete with a 2-up, open-book format and animations of page flips when you click on the right-hand page (or on the navigational arrows below). It also provides a few other types of page-view, the ability to zoom in and out on particular details, and an audio option that will read the book aloud to you (albeit in a very digital-sounding voice). Users can also embed particular books in their own websites, much like YouTube allows users to do with videos. And although all of these features except the audio option are also available in Google Books, in my personal view the Open Library interface is considerably more intuitive and attractive in its implementation of them. And for visually impaired users, of course, the importance of the audio feature cannot be overstated.

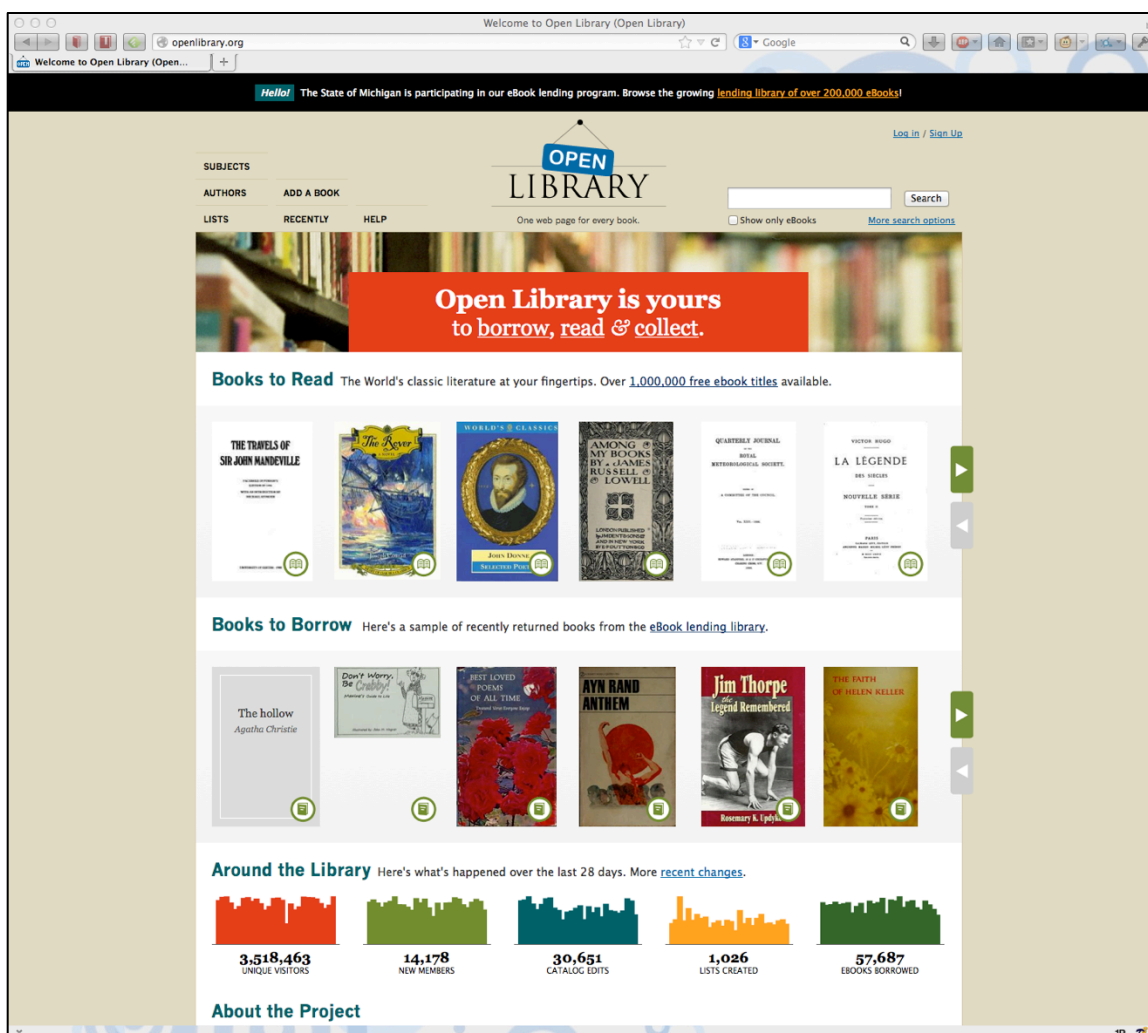


Figure 63: The Open Library Homepage

Moving back out from the book viewer, Figure 63 shows the homepage of the Open Library site. Gone are the animated gifs, tiny text, and strange information architecture of the Internet Archive site, replaced by a useful division of books into items one can read online (i.e. public domain books) and items one can borrow online (i.e. in-copyright books), presented through an uncluttered, easily-scanned interface. The site also recognizes the user's location, and offers up services specific to his or her region; in my case, this is shown by the black banner across the top of the page, alerting me to the availability of 200,000 Open Library books for lending in the state of Michigan.

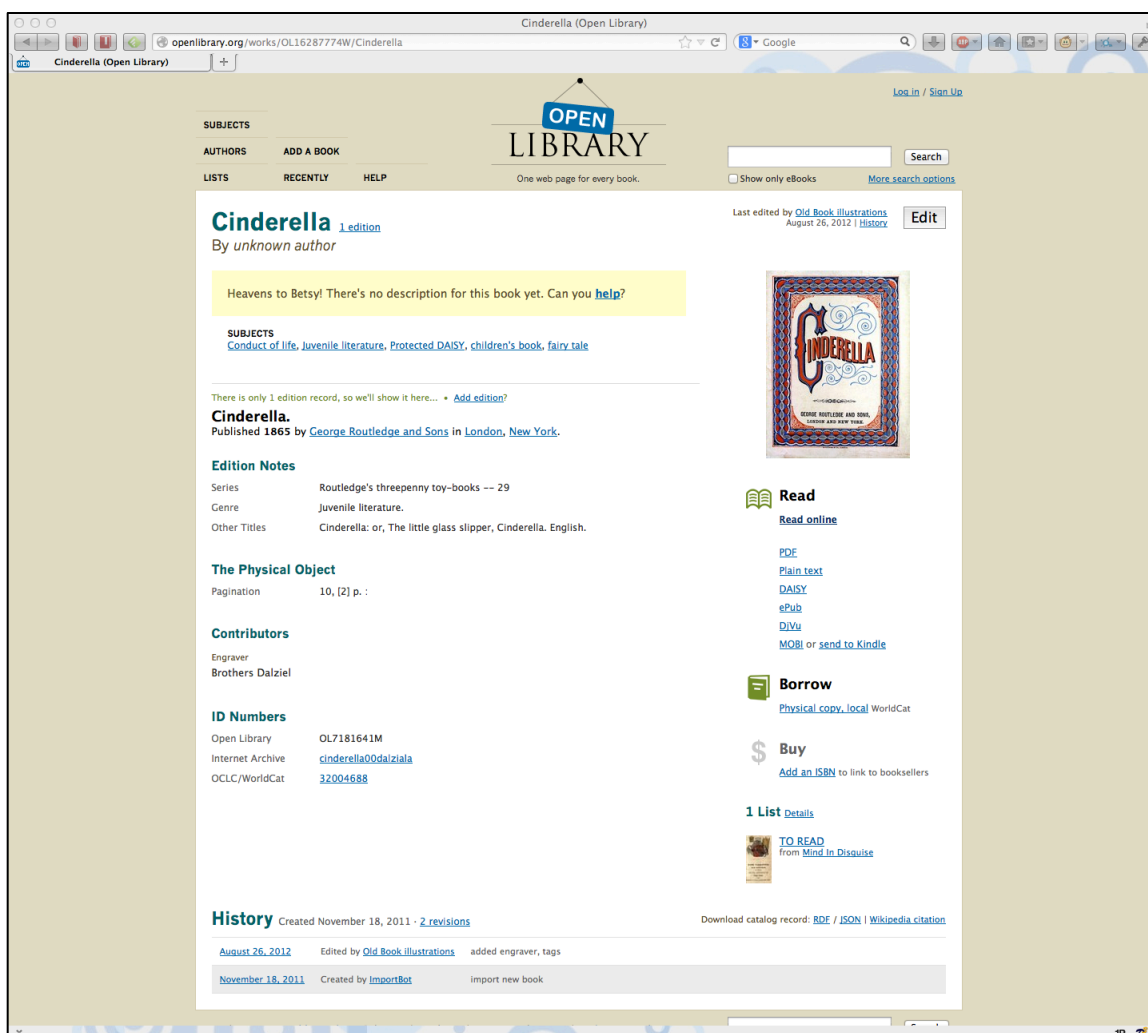


Figure 64: Item record for the same edition of *Cinderella* on Open Library

The item records on the Open Library site (as in Figure 64) tend to contain much the same metadata as the item records on the IA side, but presented more cleanly and without the user reviews (or, again, the animated gifs). However, the most interesting feature of the page depicted above – and all the others like it – has less to do with what is on the page, and more to do with what could be placed there, and how. That is, all of the Open Library records are Wiki-editable by anyone who signs up for a free Open Library user account; indeed, many contain somewhat cheeky invitations to edit them (the one above, for example, declares, “Heavens to Betsy! There’s no description for this book yet. Can you help?” – with the word “help” linking to the “Edit” page for the record). Should a user decide to go ahead and edit the record, the site presents the data entry task as a series of questions to make the fields easier to interpret. And for more advanced users, there is even a “Librarian Mode” which opens up several more fields for editing, such as copyright date, series affiliations, and formal classification information. Both the standard and “librarian” interfaces are shown side by side in Figure 65, below.

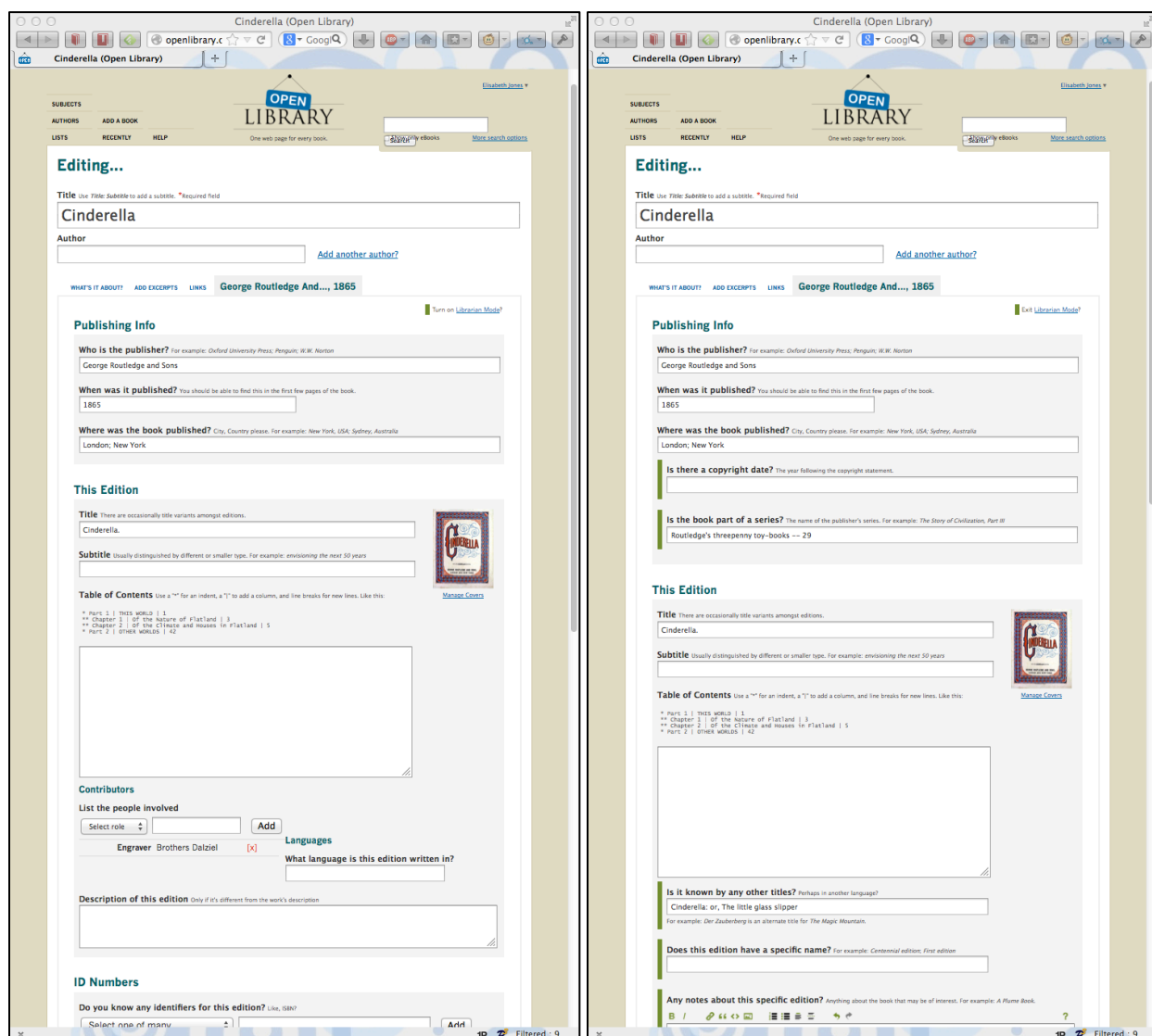


Figure 65: Standard editing interface (left) and “librarian” editing interface (right) for *Cinderella* on Open Library

In different ways, both the Internet Archive interface and the Open Library interface embody the OCA motivations and definitions discussed above. Both sites’ embrace of user-generated and user-contributed content, as well as their provision of a multiplicity of access methods and formats, solidify the OCA and IA commitment to openness in technological form. The IA site’s organization of the collection by source library places the multiplicity and multivocality of the project, and the full empowerment of each of the institutions listed, at the heart of its navigation structure (despite the relative inutility of this arrangement for search and retrieval by users), signaling the importance of at least the appearance of an egalitarian power structure among contributing institutions and the IA. And both the wiki structure of the Open Library site and, in a way, the outdated anti-usability of the IA interface seem to beckon users to become contributors; to become a part of Kahle’s “Wikipedia generation,” not content to merely consume, but thirsty to create; to fill in the

metadata boxes and flesh out the Open Library's index, or to scrape the IA's content and build a new, improved, interface of their own (or better still, dozens of interfaces, or hundreds). The IA and Open Library interfaces represent open jumping-off points rather than closed, polished systems; spaces for building rather than spaces for merely consuming. Although they might not always make for the most attractive visuals or the most usable navigation systems, these features are deeply in sync with the principles and definitions laid out by the IA and its partner libraries during the time of the Open Content Alliance, and provide a valuable alternative to the myriad consumption-oriented options already available.

4.2.2 Platform Approach

For Kahle and for the OCA, the ideal of "openness" was frequently closely linked to the idea of providing an "open data platform" – a pool of data available to be drawn upon by anyone to build whatever it is they might want to build. Indeed, within the coding of the primary sources for this case, the theme of "openness" co-occurred with the theme of "platforms" nine times, out of a total of 20 and 15 instances of each individual term respectively (specifically, these themes co-occur in P1, P2, P5, P11, "Internet Archive Receives Grant" 2006, Kahle 2007b, Kahle 2011). The close tie between these themes in this case is conceptually significant on a number of levels.

First, developing a platform rather than a portal served as one more point of differentiation between the OCA and Google Books, in keeping with the OCA's overarching goal of providing an alternative to that project. Google, as already described, primarily provides access to its book scans through a single, centralized portal, subject to various constraints on reuse. The idea of opening its Books database as a platform for the creation of a multiplicity of competing portals would be anathema to Google's approach, which restricts activities like bulk downloading and re-hosting specifically so that readers will use *Google* to execute their book searches and not any other service. After all, tracking those searches (and so many others) is how Google gets all its user data, and thereby its advertising revenues. The OCA, by contrast, explicitly set out to make all its resources available for anyone to use as a basis for building their own services; to cater to the "Makers" described above as well as to mere consumers. As Kahle described during the collaboration's heyday,

Like Wikipedia, the data [on Open Library] is editable and bulk downloadable. We see this as crucial to build a solid framework that can grow and evolve as well as inviting other non-commercial efforts to leverage the collection. (Kahle 2007b).

And an interview participant extends this assessment to the Internet Archive's content more generally, noting that

Anybody can skin it, you know, all the stuff in the Internet Archive, you can deep link,

you can embed, anybody can skin it, there's no restrictions, um...uh...and some people've done, there's an alternative interface to the music collection for example, done by some guy, who wanted to create a better one (P1).

Very much like the open source approach to building software (a link explicitly made by Kahle on more than one occasion (e.g., 2005c, 2005a)), the OCA approach to digitization was to release the book scans as a sort of content-layer source code, open for anyone with a better idea to tweak, remix, and build upon.

A second useful feature of this approach is that it avoids what one participant refers to as “the pretty portal problem,” wherein project teams expend all their time and effort building a single, attractive interface to a set of content (often to attract or appease funders desiring “shiny things”), rather than doing the less sexy work of making the backend extensible, scalable, and interoperable. As she puts it,

we [in this school of thought] think that it's important to design information so it can appear in any kind of discovery environment. So that you have some kind of open platform that's being picked up in searches and in... Flickr and in... all kinds of things that people are using. Where people live. You would have to push content out (P11).

The platform approach provides a long-term, adaptable service, to which many portals can be built serving many different end-user needs; it does not pre-judge or pre-restrict its user base, and so remains much more neutral and apparently democratic than a “one-portal-to-rule-them-all” approach like Google’s,²⁴² and more in keeping with the visions – or anti-visions – of the user prevalent among OCA partners. It also makes a much more efficient use of limited digitization funding, in that the platform has the potential to be used in many more ways and to remain useful over a longer period of time than any single portal could (given the rapid flow of technological innovation and obsolescence).

Still, the platform approach is not without its problems. A third implication of this strategy, at least in the context of the OCA, is that it can allow the digitizing parties – those with the most passion for the project – to abdicate responsibility for providing effective end-user services, beyond just APIs. As noted above, the only interface through which access to OCA scans has consistently been offered since 2005 is the Internet Archive Ebook and Texts Archive site – and that interface leaves much to be desired from a usability perspective. And the interfaces promised by Microsoft, Yahoo, and “the Open Content Alliance” (when it looked like it might be an independent organization) all failed either to persist or to ever materialize in the first place. The Open Library site is a considerable improvement in this regard, and a promising sign for the future; and yet, its existence and promise does not make the neglect of effective user interface design in all of these other cases, more

²⁴² Though of course Google’s library partners could make – and have made – their own interfaces to Google-scanned content, the Google interface is undoubtedly far and away the most broadly used.

specifically tied to the Open Content Alliance, any less disappointing.

4.2.3 Standards

One might expect that a key element of an initiative taking a platform approach to digitization would be the establishment of a core set of consistent standards, especially for metadata and imaging. After all, if the idea is to create a pool of content for others to build upon, ensuring consistency within that pool would greatly facilitate the desired reuse. However, on this issue, as in the area of collection development, the OCA's commitment to full enfranchisement of every participating institution (as well as its relatedly diffuse distribution of power) tended to interfere with its ability to reach consensus on issues of collective interest. Thus, although the scanning conducted through the Internet Archive's Scribe system did (and still does) follow a common set of standards and procedures (Miller 2012), not all of the OCA scanning passed through that system, and thus not all was subject to those standards. Instead, many libraries opted to scan some or all of their materials in other ways, using other technologies, that produced output with varying parameters (e.g., Hadro 2009, Kahle 2010a, Biodiversity Heritage Library). And alongside the OCA, of course, the IA has always welcomed pretty much any kind of scan from any source, simply advising potential uploaders to stick to one of several preferred file types (PDF, JPEG2000, JPG, or TIFF), and noting that,

for many materials, adequate images can be made with off-the-shelf scanners or good-quality digital cameras. For best results, use the highest resolution your device is capable of. Most images we process were produced at a resolution of 300-600 ppi (Internet Archive n.d.-d).

Although admirably open and egalitarian, these policies are not exactly a recipe for consistency in imaging, and can make rough work for large-scale secondary users, as in the aforementioned case of HathiTrust, where the diversity of applied standards results in a great deal of downstream work re-validating the data as it flows in (P8).

Whatever the technological inconsistencies amongst the OCA's and the IA's scans,



Figure 66: Plate from L. Frank Baum's *The Wonderful Wizard of Oz*, on Archive.org

however, it should be noted that most are nevertheless of excellent visual quality – full color as opposed to Google’s (mostly) black and white, blissfully free of scanner-fingers,²⁴³ and reproduced at high resolution. Figure 66 shows one typical page image to be found on the Internet Archive’s Ebook and Texts site – a lovely, clear, full-color image from *The Wonderful Wizard of Oz*.²⁴⁴ With regard to character recognition, the Internet Archive does not do especially much better than Google, as at least one participant noted (P15), but its OCR output is sufficient for most purposes, and unlike Google’s, it is downloadable with each PDF.²⁴⁵ Despite the diversity of workflows and technological formatting issues, it seems clear from working with these end-products that all involved shared (and continue to share) a commitment to imaging quality.²⁴⁶

With regard to metadata, there is some indication that the Internet Archive’s practices may be subject to some of the same criticisms as Google’s, particularly with respect to subject headings and record truncation. Like the Google partners, the OCA partners sent bibliographic records along with the books they were having scanned, and like Google, the Internet Archive would alter and/or truncate those records before sending back or posting the digital copies. As one OCA librarian recalled,

we would send MARC records or MODS records and we would get some kind of truncated... we used to call it Brewster Core! [laughs] I don't even know what it was. But we would get Brewster Core back, and uh, you know it was... it was an Archive-used number of fields that worked in their database. And so for the Archive's content to be truly interoperable you know there's a whole world of metadata work that probably has to happen (P11).

The participant who noted HathiTrust’s issues with OCA quality control also suggested there may be problems with IA’s metadata, pointing to the continuing frequency of manual metadata entry at IA, and the higher error rate that goes along with that (P8). And within the IA interface, the combination of this sparse metadata with an under-powered, non-full-text-capable search feature makes for some very challenging navigation indeed.

Still, once again, the Open Library site provides reason for optimism. As noted, Open Library encourages users to add to, revise, and improve any and all metadata on the site,

²⁴³ Given the configuration of the Scribe technology, one would guess that to leave a finger on the page might be to have that finger painfully crushed, since unlike Google’s process, the Scribe uses a triangular piece of platen glass to press the pages flat for the camera. As such, the absence of fingers on IA scans is perhaps fairly unsurprising.

²⁴⁴ This page is accessible at: <http://archive.org/stream/wonderfulwizarado00baumiala#page/20/mode/2up>

²⁴⁵ I have often had occasion to appreciate this feature in gathering documents for the historical pieces of my research, as being able to download the OCR saves me the trouble of running ABBYY FineReader (the same software used by IA) over the document myself.

²⁴⁶ Though, interestingly, as Leetaru notes, this commitment to quality can get in the way of access to these materials in already-underserved areas, where bandwidth is scarce. After all, with high quality comes large file sizes, and it can take a great deal of time for large files to download or to load in a browser over, for example, a dial-up connection (2008).

and does not limit those additions by imposing any of the constraints typical of library cataloging, such as restrictions on the number of subject headings that can be added. And both my interview with an Open Library designer and my own interactions with the site tend to indicate that this approach is beginning to reap some rewards. As the designer exclaimed, “*People love typing stuff in boxes on the internet!*” (P7) – and the Open Library system capitalizes on that collective enthusiasm. Although many of the records remain untouched, and are thus equally as sparse as those on Archive.org, the search function on Open Library seems to work much better, there are useful linkages provided between different editions of many works,²⁴⁷ and commonsense meta-categories are provided to group different types of tags – e.g. “Subjects,” “Places,” “People,” and “Time.”²⁴⁸ The Open Library’s metadata system remains far from perfect, but in contrast to both the IA site and the Google Books site, it does at least provide users with a means to help improve it (that is, if you see something wrong, fix it!) – and notably, that means of improvement is deeply consistent with the principle of openness that has underlain so many pieces of the OCA and its successor projects already discussed.

5. Conclusion

When the Open Content Alliance was first announced in 2005, it seemed like the beginning of a new era – in some ways even more so than after the announcement of Google Books. After all, when only Google was engaging in book digitization at mass scale, it seemed like it could be a fluke; a passion project or a whim of someone within the corporation, which could disappear at any moment. It felt inspiring, yet tenuous. When the Open Content Alliance came on the scene, however, it felt like Google’s project had expanded into a broader social movement; like the collective will had gotten behind large-scale book-scanning, and that scanning was ever-more sure to happen.

However, as this chapter has described, those hopes were not ultimately realized through the Open Content Alliance itself. For a time, the OCA managed to mobilize a large and diverse alliance of institutions with complementary motivations around a common set of principles, and succeeded, for a time, in translating that mobilization into practical results: over 1 million books scanned and made available through the Internet Archive from 2005-2008. But then it lost its way. Conflicts over how the Alliance should be governed, coupled with the loss of major funders, led the OCA to dissolve so completely that some of its most influential leaders now claim that it never existed in the first place, or that it was only ever a

²⁴⁷ See, e.g., the Open Library page for *An Essay on the Demoniacs of the New Testament*, which lists five editions, two of which are available to read in full online (http://openlibrary.org/works/OL248315W/An_essay_on_the_demoniacs_of_the_New_Testament). (Also, fun fact: one of the two full text editions was scanned from the personal library of U.S. President John Adams, now housed at the Boston Public Library.)

²⁴⁸ See, e.g., the Open Library page for Charles Dickens, http://openlibrary.org/authors/OL24638A/Charles_Dickens

set of principles, and never a functioning entity (P3, Kahle 2011). And though I do not doubt that this assessment reflects these individuals' honest view of the situation, I also disagree with them. The Open Content Alliance *was real*; it just gradually ceased to be so somewhere along the way. And although some early OCA contributors might take exception to this assessment,²⁴⁹ I concur with the view of one librarian interviewed, who contended that the OCA had

planted the seed for more regional collaboratives, [which are being pushed] now under DPLA. And I think HathiTrust is its spiritual successor. You know, even though Hathi was probably in the works very very soon after OCA started, um... It's definitely, like I say, I see it as the spiritual successor of what, of what OCA was starting to be (P11).

As noted in the introduction, the leaders of the OCA dropped the baton, and it shattered – but its pieces live on. The scanning infrastructures and partnerships developed under the OCA paved the way for their continuation through the DPLA and through more local initiatives. The body of actual scans the OCA produced have helped to seed the HathiTrust, which itself was founded as a collaborative endeavor among libraries, undoubtedly drawing on many of the lessons learned through collaboration under the OCA, which shared many of the same partner institutions. And the principles of openness in scanning championed under the OCA, however problematically implemented under that initiative, continue to ring through in the ongoing scanning efforts of the Internet Archive, as well as its Open Library project, and also in the DPLA.

This case, as is likely quite clear, has been a difficult one to synthesize. The narrative threads often meander and fail to cohere; the data and even the views of reality behind the data often conflict and even contradict one another; and the edges of the case blur and fade into others at several points. But on a certain level, all of that makes perfect sense, as that is the reality of the OCA. Tangled, contradictory, blurry, faded. It lacks the tidy linear history of the Boston Public Library, or even of Google Books, as well as the relatively neat conclusion of the Carnegie Library Program. Indeed, unlike that program, the Open Content Alliance was never even officially canceled – those involved just quietly stopped enacting its existence. It is the case that wasn't a case, except when it was. A story of conflicting realities, or many different stories, or perhaps not a story at all. An inspiring set of achievements, an incubator for impressive successors, and a disappointing failure, all at the same time. The Open Content Alliance was both less and more than a challenge to Google – that is, assuming you believe, as I do, that the Open Content Alliance ever *was*.

²⁴⁹ Especially Brewster Kahle, who repeatedly equated HathiTrust to Elsevier (a common bogeyman in the library world, known for its extortionate pricing and iron-fisted access controls) throughout my conversation with him.

Conclusion

The four case chapters above have provided a significant amount of in-depth analysis of each case. This final chapter will draw connecting lines between these analyses, identifying common themes as well as divergences between cases. To preface this synthesis, however, it is worth noting that these cases are only partially separable in terms of motivation, definitions, or implementations. After all, none of these cases happened in a vacuum; rather, later cases built on earlier ones, while contemporaneous cases learned from and played off of one another.

Structurally, this chapter will be largely consistent with the previous four, exploring each of the three research questions – the why, the what, and the how – in turn. It will then conclude with a discussion of the limitations of the research and a brief meditation on the broader issues surrounding the strivings toward informational utopia it has explored.

1. Motivations

Assessing the motivations of a person or group is inherently one of the most fraught exercises in social science. The social scientist cannot enter the subject's head, and thus cannot truly know what initially compelled them to action, or whether subsequent events might have colored or even radically altered their perception of that initial compulsion upon later report. The most we can do is sketch around the edges, piecing together things said at the time, later reflections on events, and the external impressions of others also present to shade in an imperfect image of the "why" behind the "what" being examined. Indeed, as noted in the introduction to this dissertation, even if we allow that individuals make decisions and take actions based on a constant reflexive monitoring of their conditions and their own prior actions (Giddens 1984, 5-6), the specific motivations behind any given action tend to remain out of sight, possibly even to the actor in question. And particularly in research drawing predominantly on historical materials and the memories of individual interviewees, many factors can intervene between the actual motivation as it was felt at the time and the motivation as interpreted later: for example, after-the-fact rationalizations by the individuals involved, divergences between the lived experience of the motivation and what was written down about it at the time (or later), bias introduced in the portrayal of motivations depending on the audience for the portrayal – all of these and more can complicate understanding of what "actually happened." Indeed, at some point one may even begin to question whether there is any objective truth – any "actual" – to be found in this arena. The best we can do is assemble as many points of evidence as possible, and sift among them for grains of consistency, despite whatever flaws and biases any one source might exhibit.

Still, also as noted in the introduction, any analysis of what Giddens calls "strategic conduct," wherein actors draw upon structural features to construct new socio-structural

arrangements, rather than simply perpetuating and reconstituting existing conditions (1984, 288-89), requires some account of motivations in order to understand the new configurations thus created. In this dissertation, the focus has been precisely on such strategic conduct, and has thus specifically sought to explore the motivations behind the activities of particular actors – individuals with some capacity to shape a particular library system or LSDI – as reflective of and embedded within particular social settings and value systems. And further, on a practical level, to the extent that an understanding of the motivations of various actors within these projects has been developed, that understanding forms an illuminating backdrop against which to view the steps taken by those actors in bringing the projects to fruition. Though these agents’ purposive actions often led in other directions than were originally intended – as, indeed, Giddens claims is inevitable (e.g., 1984, 294) – teasing out some of the motivations behind them helps to describe at least the beginning of their trajectories. This aim is encapsulated in the first research question posed at the outset, reiterated below:

RQ1: In the four large-scale IDEs under consideration, what were the initial motivations for pursuing each project expressed by and among key stakeholders (those with some capacity to influence what happened)?

The answers to this question for each case have been presented in the four case chapters above. At this point, however, it will be illuminating to examine the through-lines and discontinuities among the cases. As a starting point for drawing these connections, the table below presents the major subdivisions of the motivational piece of each case chapter:

Boston	Carnegie	GBLP	OCA
• (Adult) Education • Egalitarianism & Uplift • Moral/Character Development • Municipal Prestige/Competitiveness (secondary) • Economic Development (secondary) • Public Demand (secondary)	• Carnegie himself: ◦ Wealth as virtue, virtue as duty ◦ Nothing for nothing ◦ Evolution, not revolution ◦ Ladder for aspiring ◦ Personal History/Affection for Books ◦ Shifting Motivations over time • Communities: ◦ Local pride ◦ Education ◦ Social Palliative	• Google: ◦ Mythos/mission ◦ Profit motive (or not) ◦ Get more data • Libraries: ◦ Mission (broaden access, preserve information) ◦ Pragmatic Aims (Support/extend library services, Improve library rep/relevance)	• Brewster ◦ The Collector ◦ The Tech Entrepreneur ◦ The OA Visionary • Principles ◦ Anti-Googlism ◦ Openness ◦ Procedural Transparency ◦ Preservation • Pragmatics ◦ Improve technologies ◦ Keep up with tech ◦ Cost Reduction

Within the above table – as well as within the chapters it represents – several motivational continuities and divergences emerge. However, before delving into these comparisons, it is necessary to tease apart a significant meta-issue within both the research question and the thematic divisions above: that is, the question calls for the exploration of initial motivations

among “key stakeholders” – “those with some capacity to influence what happened” – and in the table above, it becomes clear that in most cases, more than one such significant stakeholder group existed (hence the major divisions within the lists for all but the Boston case). As such, it will be useful to dwell for a moment on exactly which stakeholder groups were “key” in each case, and where their particular spheres of influence lay.

1.1 Positioning “Key Stakeholders” Across Cases

The question of the differential relevance of various stakeholder groups involved in the development of a technological system (to borrow the terminology of SCOT theory) is, at its core, a question of power relations. In exploring the imbalanced constellation of agency and motivation among the stakeholder groups in these four cases, I draw upon the basic definition of power offered by Giddens. That is,

Power is the capacity to achieve outcomes....Power is not, as such, an obstacle to freedom or emancipation but is their very medium – although it would be foolish, of course, to ignore its constraining properties. The existence of power presumes structures of domination whereby power that ‘flows smoothly’ in processes of social reproduction (and is, as it were, ‘unseen’) operates (Giddens 1984, 257).

In essence, within social life, various processes are continually carried out by agents, subject to particular structures of domination, which serve to reproduce (and/or transform) the fabric of social life. Within these processes of social reproduction, “having power” can thus be understood, on some level, as having the ability to enact one’s will, given the social context in which one acts. Those in positions of privilege within existing structures of domination can generally marshal resources – money, supplies, other people – that further enable the achievement of their desired outcomes. Within the cases examined here, various stakeholders wielded varying levels of power in striving toward the realization of their specific motivations. Some were in a position to directly affect the shape of an institution or project, while many others saw their motivations and desires mediated through other stakeholder groups in ways that blunted their impact on what happened (though notably, a blunted impact is an impact nonetheless). The following sections will briefly sketch the power structures present among relevant stakeholder groups in each of the cases examined here, with a particular focus on positioning the voice of “the public” in each case.

1.1.1 The Boston Public Library

In the early years of the Boston Public Library, two groups were primarily responsible for shaping the institution’s development: the Boston City Council and the Library’s Trustees. These two groups were not unconnected: the Trustees served at the pleasure of the City Council, and there was overlap in their membership. However, the Trustees’ relationship with the City Council was often adversarial. The City Council had final say over all library funding and personnel appointments (at least until 1878, when the Trustees were

incorporated as a more independent entity (Wadlin 1911, 76)), while the Trustees were more intimately involved with developing the vision for the library and, eventually, with operating it. For the most part the City Council seems to have accepted the reports and recommendations of the Trustees without amendment, but when clashes did occur, it could become quite destructive – as, for example, when such a disagreement caused Justin Winsor to leave the library's Superintendency.

Despite the City Council's early hold on the purse strings, however, the Trustees were quite an influential group in the BPL's early history. And, as discussed at length in the chapter on this case, two Trustees – George Ticknor and Edward Everett – were especially prominent. These two men largely defined the vision for the institution between them, with some input from the institution's major donor, Joshua Bates (whose influence, though significant, was virtually entirely mediated through the Trustees with whom he corresponded, as he lived in London and never even personally visited the institution his donations helped to found). Ticknor in particular, moreover, continued to maintain direct involvement in the running of the institution until his death. Under the leadership of Ticknor and Everett, the Trustees were responsible for virtually all of the shaping of the BPL that occurred prior to its actual establishment in 1854 – the development of policies, collection goals, physical storage and retrieval arrangements, and so on. Most of the motivations that were relevant to the establishment of the institution, and that have accordingly been discussed here, were the motivations expressed by the Trustees, and especially Ticknor and Everett.

Once the library was established, the library staff, and especially two of the early Superintendents – C.C. Jewett and Justin Winsor – began to have a strong role in shaping the institution on the ground – arguably a stronger role than the Trustees, as the staff were present on a day-to-day basis, enacting policies and providing services. At that point, the Trustees' views became mediated through the staff (with the Superintendent largely mediating between the Trustees and the rest of the staff), while the City Council's views were mediated through the Trustees (and then on through the Superintendent to the staff).

The public (i.e. the citizenry of Boston) then appears at either end of this chain of mediation: it is their views that are theoretically being represented by the actions of the (elected) City Council, which flow through the Trustees to the Superintendent and along to the library staff, but at the same time, they are also the library's users, in direct contact with the institution itself as well as its staff at the point of consumption. At neither end of this spectrum does the public wield much effective power, however. As users, they are subject to the rules, norms, and physical structures in place at the library, without much recourse for directly altering those rules, norms, or structures. And as voters, their opinions and agendas are mediated through many levels of other stakeholders, each with opinions and agendas of their own. And further, those unable to vote – women, children, recent immigrants, people of color – stand at yet another level of remove, as their voices must be

heard and passed along by those who do possess the franchise if they are to have any influence at all in that direction. Notably, these non-voting demographics likely had more influence as library users than they did as citizens (or residents): as users, they could have direct contact with the staff, and although the staff may not have had much say in the institution's formal policies or architecture, they likely had a fair bit of power in determining the extent to which those formal policies would be followed and enforced on the ground, and it is plausible that they may have taken user input into account in such determinations.

1.1.2 The Carnegie Library Program

In the Carnegie Library program, it could superficially appear that all the power rested with Carnegie, or at least with Carnegie via Bertram. This is not the case, however.

Carnegie himself was certainly the primary – if not the solitary – mover in bringing the program as a whole into existence. Without his personal motivations and capacity to enact them, no other part of the program – including the power differentials between the central administration of the program and the requesting communities – would ever have existed. As such, Carnegie's personal motivations absolutely mattered in shaping the program as a whole. And at the point of accepting or rejecting each request for library building funds, Carnegie and Bertram wielded virtually all the power; at that specific point, they could and did dictate who could make requests, what information had to be provided, what evidence was necessary to illustrate the community's need, and, later, how the building they funded would be designed and built.

Still, the lion's share of the agency brought to bear in (a) marshaling support for and submitting these library requests in the first place, and (b) actually constructing, filling, and running the library once the donation had been granted, rested within the local communities themselves. Given this, although the local communities had no control over how the program of donations as a whole would be run, they had a great deal of control over how their local iteration of a Carnegie Library would play out – and Carnegie and Bertram had virtually (or perhaps actually) none. There was a great deal of diversity among the communities requesting Carnegie donations in terms of motivations, political realities, and distributions of power, and as such, it would be impossible to capture all the nuances of those diverse contexts here. However, a few general points are worth noting. The voices of local women were often quite relevant in obtaining support for requesting Carnegie libraries, and for gathering the initial local funding for those libraries. Through these activities, these women's groups made their motivations matter in shaping their local institutions. At the same time, however, as in Boston, women did not yet have the right to vote or hold office during the period of Carnegie library donations, and so, as detailed in the Carnegie case chapter above, their voices had to be mediated vis-à-vis Carnegie and Bertram, via the local elected leadership (and those with the capacity to elect that leadership

- namely, the local men). And of course, as also noted in Chapter 4, nonwhite individuals, and especially African Americans in the South, also lacked the franchise, and were still less likely to gain the ear of the local white officialdom than were the local (white) women's groups (who, though lacking the vote, undoubtedly included many wives, daughters, and mothers of those local officials, and thus likely had some personal sway). Additionally, many of the issues with mediation regarding ongoing funding were likely similar to those in the Boston case: local officialdom was responsible for levying taxes and doling out public funding, but that local officialdom was only directly accountable to those with the right to vote - not the entire community. This would have affected both the possibility of making a Carnegie request - since proof of ongoing tax support was required for all such requests - and also the ongoing survival and priorities of the library, depending how much funding was actually made available, and with what strings attached.

1.1.3 The Google Books Library Project

The power dynamics among the partners who were involved in the Google Books Library Project before the project's public announcement were quite different from the power dynamics that would obtain for partners who joined the project later. Specifically, the first five, or perhaps six, libraries to sign on with Google had much more say in how those partnerships would progress than later library partners would have. Although Google was the dominant partner even in those early partnerships (not least because it was the one bringing the giant bags of cash), the company initially needed to woo the libraries a bit. After all, Google was proposing a venture that seemed like a pretty crazy idea at the time, and they needed those early, elite libraries' participation for it to even begin to work, on multiple levels: namely, gathering the content, gaining the imprimatur of institutions whose record of custodianship and service in the book domain was far better established than Google's own, and figuring out the procedures through which the ambitious initiative would actually be run. On this last point, as previously noted, librarians from three of the first five partner libraries - Michigan, Stanford, and Harvard - were especially crucial, and for that and other reasons, those institutions were able to make their voices heard more forcefully than any other libraries would as the project progressed. Still, even among these three, Michigan and Stanford seem to have become more influential than Harvard, mainly because they were more simpatico with Google's aims and procedures, but also likely due to the personal involvement of Larry Page, an alumnus of both institutions, in garnering their support. Indeed, in part due to a top-level leadership change at the Harvard Libraries, from a supporter of the GBLP (Sidney Verba) to a prominent critic of the project (Robert Darnton), Harvard backed away from the project after only a few years, having scanned about 1 million public domain items with Google - and from that point on, they put all their weight behind criticizing the project and envisioning alternatives to it, most prominently the DPLA. The NYPL and Oxford's Bodleian Library seem to have been more

“along for the ride” in these early years, but as founding partners, they were also likely granted more influence than later libraries at least over the content of their particular contracts; however, that assertion is difficult to confirm, since neither contract has been made public.

In all five of these early partnerships, Google was willing to give the libraries more of what they wanted just to get them on board, and were also more naïve about the long-term implications of the promises they were making – for example, with regard to scanning entire libraries. (This naiveté had been cured by the time the later, more standardized contracts were drawn up.) The University of California, Google’s sixth partner, also had some degree more sway than later partners would, due largely to its size and its public status (which gives sovereign immunity, which acts as a safeguard against copyright and other litigation) – though still not as much as the first five, and certainly not as much as Michigan or Stanford. Libraries who signed on after the University of California generally had far less clout in negotiating their contracts; by that point, Google was pretty sure they could get just about anyone to sign on, and so it became something they could be selective about, rather than worrying about how well their agreements met any specific library’s individualized needs.

As far as mediated influence goes, there are a few parties whose views mattered. One is copyright holders, whose inevitable lawsuit could have been strengthened (or avoided) had Google made other decisions with regard to what it would scan and how it would use those scans. So in that sense (and rather ironically), copyright holders’ interests were mediated through Google and its lawyers and designers. There are also two groups of users involved here, with arguably fairly different interests: Google’s users and the libraries’ users. That is mainly a topic for the next major section here (on users), but at this point it’s worth noting that neither type of user had much, if any, direct impact on how Google Books proceeded (with the possible exception of users with a public podium, like Siva Vaidhyanathan, Pamela Samuelson, or Lawrence Lessig). Generally, the interests of users only had any impact to the extent that they were recognized and advocated for by either Google or an influential library partner.

1.1.4 The Open Content Alliance

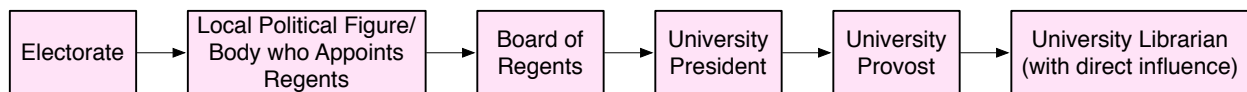
The distribution of power in the OCA, much like that in Google Books, rested mainly with a technology organization on the one hand – in the OCA’s case, the Internet Archive – and with a group of libraries on the other hand. However, beyond that superficial level of similarity, the power dynamics differed a great deal between the two cases. The OCA was initially shaped almost entirely by a single individual – Brewster Kahle. Its early partner libraries were also at least nominally empowered, but from the interviews there is a strong sense that at least some (e.g., University of Toronto, Biodiversity Heritage Library, Boston

Public Library) signed on largely because they agreed with Brewster's principles; in a sense they contributed their agency to further empowering his interests, rather than distinguishing their own – at least at first. One of those principles, after all, was a sense of egalitarianism and democracy; that all the partners should have an equal voice. And so although the partners agreed on the principles, many went entirely their own way on the practical implementation – and those implementations tended to be targeted more at serving each library's local interests and motivations than toward serving the needs of the collective whole.

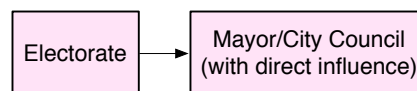
The mediated influences on the OCA were virtually identical to those in Google Books: copyright holders (whose influence was felt via collection management and design decisions) and the user bases of both IA and its library partners. And just as in Google Books, users' motivations only mattered to the shaping of the project to the extent that those motivations were recognized and supported by IA or the librarians. And notably, in neither case is there much evidence that the user base was actively assessed or consulted in any systematic way prior to making decisions about the directions of these projects; a great deal depended on the existing knowledge and/or assumptions of the designers, librarians, and businesspeople involved – and within the interviews conducted for this dissertation, it became fairly clear that many of these individuals had not given much specific thought to their visions of the end-user at all.



One notable distinction that emerges between the public library and digitization cases in these depictions lies in the role of the public in each. That is, in the public library cases, some level of public support was actually necessary in order to begin, though that support was strongly mediated through elected officials. The digitization cases required no such public imprimatur in advance of initiation (or really, at any point). The electorate cannot vote out the CEO of Google, or the head of the Internet Archive, or their local university librarian. Even in the case of public institutions participating in LSDIs, the chain of mediation is sufficiently long as to render the public's agency virtually nonexistent – something like:



In the public library cases, the chain was much shorter – more like:



In this latter diagram, many more people fall outside the box marked “Electorate” – women, people of color, etc. – and yet those who fall within the box are afforded a

considerably more direct influence on the information democratization effort du jour for the 19th century than those within the equivalent box for LSDIs above it.

Still, as noted above for the Boston case, the public has the potential to be involved at both ends of the mediation chain: they are both voters with the capacity to influence who gets elected *and also* actual or potential users of the resources created by these initiatives. Taking this into consideration, it seems quite clear that the public has much more potential for influence as a *user* of Google Books or the Internet Archive (or HathiTrust or Open Library) than as a *voter*, given the chain of mediation just described. The position of the user is typically mediated as well: through data about their behavior that can be gathered without their knowledge (i.e., web analytics), through comments delivered through “Help” or “Feedback” links on resource websites (and then through the individuals responsible for reading and/or responding to such messages, who may take those comments no further than themselves), or through critiques delivered in a public forum (e.g. the press, a blog, a tweet, etc.), which may or may not ever be seen or heard by those in a position to act on the suggestions thus offered. The agency of the LSDI user is still highly mediated and marginalized – but it is at least somewhat less marginalized than the agency of the voting public on LSDI practice, which is so highly mediated as to be totally inconsequential.

This dissertation has been primarily concerned with the motivations that “mattered” – that is, motivations wielded by individuals or groups who had a direct voice in the decision-making processes and designs of the initiatives; as such, in summing up the continuities and divergences below, the focus will remain on those groups: Boston’s Library Trustees, City Council, and early library Superintendents; Carnegie himself (and Bertram as his surrogate) and to a lesser extent the local leadership in recipient communities; Google and its partner libraries; and the Internet Archive and its partner libraries. The mediated users, in turn, will be considered through the lens of their mediators, in section 2 (Definitions).

1.2 The Spark: Personal Passions

One intriguing commonality across all four cases is that in each one, the personal history and predilections of at least one very wealthy individual formed a necessary, though not sufficient, condition for their initiation. These personalized origin stories form one of the most mythologized pieces of each case, and serve as an emotional touchstone (and frequently a rhetorical device) that helps to personalize and promote other motivations tied to idealistic and pragmatic aims. Each of the relevant origin stories has already been discussed in the case chapters, and I will not belabor them here, but these four seem to me to be the most relevant set:

Case	Individual	Parable
Boston Public Library	Joshua Bates	In Bates's own words: "my own experience as a poor boy convinced me of the great advantage of such a library. Having no money to spend and no place to go to, not being able to pay for a fire or light in my own room, I could not pay for books, and the best way I could pass my evenings was to sit in Hastings, Etheridge, and Bliss' bookstore and read with a kindly permitted me to, and I am confident that <i>had there</i> been good warm and well lighted rooms to which we could have resorted with proper books, nearly all the youth of my acquaintance would have spent their evenings there to the improvement of their minds and morals" (Bates 1852).
Carnegie Library Program	Andrew Carnegie	In Carnegie's own words: "When I was a boy in Pittsburg, Colonel Anderson, of Allegheny,—a name I can never speak without feelings of devotional gratitude,—opened his little library of four hundred books to boys. Every Saturday afternoon he was in attendance himself at his house to exchange books. No one but he who has felt it can know the intense longing with which the arrival of Saturday was awaited, that a new book might be had. My brother and Mr. Phipps, who have been my principal business partners through life, shared with me Colonel Anderson's precious generosity, and it was when revelling in these treasures that I resolved, if ever wealth came to me, that it should be used to establish free libraries, that other poor boys might receive opportunities similar to those for which we were indebted to that noble man" (Carnegie 1889a, 689).
Google Books Library Project	Larry Page	As the Google Books History page puts it, "In the beginning, there was Google Books" (Google 2007). That is, the idea for the algorithm behind Google Search originated from Digital Library Initiative-funded work related to citation linking in books – but since books were not yet digital and the Web was, the algorithm was built for the Web. And although many accounts of this history depict Page and Brin as being equally enthusiastic about the books aspect of this citation-linking origin story, Levy (2011) in particular ties that enthusiasm uniquely to Page, who seems to have been extraordinarily passionate about the idea of scanning books, and much more instrumental than Brin in pushing forward the GBLP. For example, Page was the one personally testing scanning technologies, and also personally meeting with librarians at the "Google 5."
Open Content Alliance	Brewster Kahle	All of those interviewed about the OCA link its existence quite explicitly to Kahle's passion and persona. Kahle himself, meanwhile depicts the project as an explicitly intended culmination of 25 years of his life's work, explaining that, starting around 1980, <i>"I sort of fastened on the 'why don't we just build the library'... um... idea. And it required some things to happen first. First we needed sort of computers that could store enough, um... so I helped work on that. And then we needed a network so that you could actually get to the stuff, so I helped work on that. That was the late '90s. Then we needed software systems and a mechanism for publishers to make money, I worked on that in the early '90s. Um... by the time we sort of got the computers, the network, and the, and the publishers online, sort of '94, then I thought I could really start working on the library. Um... then... so... started with materials that were, um, in danger. So started with the... first with the web, then the next was television, um... and started working on books, um... in... 2001? Um... with the Million Books Project led by Raj Reddy's project out of Carnegie Mellon with the governments of India and China"</i> (Kahle 2011).

Interestingly, the "why" behind the passions of Larry Page and Brewster Kahle for public-access book-scanning are somewhat less clear than for Bates or Carnegie. Both refer to some level of personal history, but it is personal history at the level of "I had this great idea and I had to wait for technology to catch up so I could do it" rather than an explicit desire to pass on to others opportunities that had been given to them in their youth. It is, of course,

plausible that Page and Kahle did have some parallel childhood experience – perhaps they loved their childhood public libraries, and wanted to amplify that experience to the world through the affordances of the new global networking technologies (some of which they had personally helped to develop). It is also possible, regardless of Page and Kahle’s upbringing, that there was less of a sense that this type of desire even required justification, beyond pointing to the already taken-for-granted justification behind public libraries. We now accept without question that providing the public with rooms full of books for them to freely use is a good thing – so why should it not be a noble aim to translate that to the Internet? Whatever the case, the personal passions of Page and Kahle were at least as instrumental in launching their respective book scanning projects as were the personal passions of Bates and Carnegie in theirs. In each case, these individuals’ passionate interest in this particular sort of project inspired them to pony up a large quantity of money to make their visions real – and in each case, it was enough money to get the project off the ground, but not necessarily enough to sustain it. The passion of the wealthy formed a crucial spark for each project, but it would take more than that to kindle those sparks into fully-fledged, sustainable initiatives – and not all of these projects reached that point.

1.3 Grand Ideals: Motivations Based on Principle

There is a great deal of continuity between the broad-scale principles and ideals that motivated the public library cases examined here and those that motivated the digitization cases. And that continuity is no mere coincidence, but rather reflects the historical importance of the public library movement in shaping the American (and perhaps global) image of the social role and projected benefits of providing free, public access to information across various social, economic, and cultural divides. The idealistic motivations of the LSDI cases are in this sense direct historical descendants of the idealistic motivations of the public library cases, although the modern actors espousing them did so with widely varying levels of historical awareness.

Within the early histories of both the Boston Public Library and the Carnegie Library Program, roughly four threads of idealistic motivation coexisted: (1) education, (2) equalization of opportunity, (3) class mobility, and (4) moral elevation. These threads were highly intertwined and interdependent, as should be clear from the case chapters above. In both of these historical cases – and for many other initiatives within the U.S. Public Library Movement (e.g., Ditzion 1947, Lee 1966) – the basic idea was that providing books to the general public would equalize educational opportunity, and thus enhance class mobility and moral character, for those members of the public with sufficient self-motivation to take advantage of the opportunities thus offered. And that last piece – that individuals be willing to work for their rewards – is crucial. In neither the Boston nor the Carnegie case was there any particular value placed on helping anyone unwilling to help themselves; Carnegie’s “nothing for nothing” motto sums this philosophy up most succinctly. The

library merely equalized *opportunity* by opening a door – to education, to economic and moral betterment – it was up to the public to walk through the door with the tools and literacies necessary to take advantage of the opportunities thus offered.

Within the Google Books Library Project and the Open Content Alliance, these ideals entered the motivational picture in a much more institutionalized form – that is, as the explicit missions of all of the institutions involved: Google, the Internet Archive, and most of all, the partner libraries in both projects. For each of these information-disseminating institutions, commercial and non-commercial alike, a significant portion of the mission relates to *access*: for Google, making “the world’s information...universally accessible and useful;” for IA, “Universal access to all knowledge;” and for the libraries, the goal of connecting people with the resources they need, expressed in various forms and at multiple levels of breadth (e.g., P2, P3, P12, P13, Coleman 2006, Colvin 2006, Stanford University Libraries 2006, Barnett and Willmann 2007, Google 2007, Internet Archive n.d.-a, New York Public Library). Yet, at a fundamental level, these missions simply reframe the core idealistic motivations of the public library movement with less moralism and class-oriented reasoning: the basic idea is still to equalize access to information for the purpose of enabling self-education by the motivated – to open the door to the resources, and rely on the preparation and motivation of the user to bridge the gap between physical access to materials and gaining actual intellectual value from the information contained within them.

Interestingly, the issue of preservation forms a key divergence between the mission-related motivations of the LSDI partner libraries and the early public library cases. In their earliest years, the Boston Public Library and the Carnegie Libraries certainly sought to protect their collections in various ways, though this was likely more of a priority for Boston than for most Carnegie Libraries, simply by virtue of Boston’s larger and more valuable collection. However, the goal of providing access was far more important in motivating both initiatives than was any thought of preserving the materials or the information within them. Preservation was simply a necessary corollary to the access mission; something that needed to happen in order for the libraries to continue to provide access to the materials thus preserved. But then, the Public Library Movement as a whole was to some extent an explicit shift away from locking books up in towers where nobody could touch them (the total preservation orientation) and toward some acceptance of damage and deterioration in the name of greater access to a broader range of people – and in that context, the access orientation of both the early leadership in Boston and within the Carnegie library program makes perfect sense.

While the GBLP and the OCA were also more motivated by access provision than by preservation, preservation played a far more dominant role in the rationales given by libraries for participating in large-scale digitization than it had in the expressed motivations of the early public library projects. And it is not difficult to ascertain why. Consider: very

few of the libraries involved in the GBLP or the OCA were public libraries, and those few that were public were also research libraries (e.g., BPL, NYPL). And historically, research libraries have not been nearly as keen to privilege access over preservation as public libraries have been; rather, preservation and access within research libraries are held to be equally important goals, as research libraries generally see themselves as serving as custodians of history and culture: access in the present is no more important to them than access in the future; both must be maintained. As such, it is logical that these types of libraries should emphasize the preservation benefits of digitization in parallel to the access benefits. And those benefits are twofold: first, by digitizing, the libraries gain digital avatars of their resources that can be offered for use in place of potentially fragile paper versions; and second, digital files can be multiplied and redundantly stored in many geographic locations, which helps to ensure that the information in them will persist even if some catastrophe destroys the library or libraries in which the physical materials are held (examples of such catastrophes offered by LSDI participants include Hurricane Katrina, the destruction of libraries under the Khmer Rouge, and the eventual “Big One” that is projected to someday occur along the San Andreas fault (P2, Coleman 2006)). As such, while it may seem superficially odd that the stately (and now themselves historic) public libraries of the nineteenth century should be less motivated by preservation than the fast-moving digitization projects of the early twenty-first, on deeper inspection this divergence makes sense, as it is in keeping with the types and missions of the institutions involved in each case.

1.4 Brass Tacks: Pragmatic Motivators

The idealistic motivations behind both public libraries and large-scale digitization initiatives are well known, in no small part because those ideals make for the best rhetoric: the leadership of both types of project have been trumpeting these ideals from the rooftops from the first moment the ideas for them occurred. And yet, there is another set of motivations common to both early American public libraries and modern LSDIs which have also played powerful roles in spurring them into existence – a pragmatic set – and I would suggest that these pragmatic motivations may even be more interesting than the well-known ideals, if only because they have been much less recognized and discussed. In particular, this section will highlight two pragmatic motivations that played roles in all or most of the initiatives examined here: namely, (1) competitiveness and (2) a desire for financial gain.

1.4.1 Competitiveness

Within both of the public library projects and both of the digitization initiatives this dissertation has examined, competition played a powerful role in gathering the support of vital constituencies. In the public library cases, this competition existed among

municipalities. The necessity of founding a major public library in Boston was repeatedly justified to the Boston City Council in part by arguing that Boston could not stay competitive with New York if the city remained without a world-class public library (e.g., Seaver 1852, "Fifth Annual Report of the Trustees of the City Library" 1857). In the case of the Carnegie Libraries, a similar sort of inter-city keeping-up-with-the-Joneses occurred, but diffused over hundreds of smaller municipalities: if East Springfield had a Carnegie Library, well by gosh, *West* Springfield had better have a Carnegie Library too, to keep up with the neighbors! Particularly in the context of the late nineteenth and early twentieth centuries, when cities and towns throughout the central and western U.S. were more actively jockeying for status and recognition against their perceived peer localities, these competitive arguments frequently served as compelling calls to action on the library front.

Within the digitization cases, the issue of competition comes up in a different way, though some parallels exist. That is, within the GBLP and the OCA, the competitive spirit was most evident among the libraries signing on with one initiative or the other. On the one hand, for the Google libraries, there was a sense that signing on with Google – or perhaps more accurately, being selected for inclusion by Google – served as proof of the world-class status of their collections, their ability to stay on the cutting edge of digital technologies, and, on some level, their basic “coolness” (“Stanford and GBS: Statement of Support” 2005, Carr 2005a, Coleman 2006, Colvin 2006, Keio University 2007, e.g., Committee on Institutional Cooperation 2008, Cornell University Library n.d., University of Wisconsin Libraries n.d.-a). For the Open Content Alliance libraries, on the other hand, joining up was also a signifier of excellence, but along a somewhat different set of parameters. There too, signing on was taken to demonstrate a commitment to stay on the cutting edge of technologies, but beyond that, it was also seen as signifying that their institutions were more deliberate and respectful of existing institutions and processes, and that they occupied the moral high ground relative to the libraries that had signed on with Google (e.g. P2, P11, “Boston Library Consortium and Open Content Alliance to Provide Digitized Books” 2007). And of course this latter piece – the moral high ground – came part and parcel with the other level on which the OCA relied on competition to win supporters: its fundamental opposition to and competition with Google, and desire to provide a “better” – more legal, more open, more transparent – alternative to what Google was doing.²⁵⁰

In all four cases, notably, we see cities and institutions striving to compete with one another on the basis of who can make the most information available to the public within their

²⁵⁰ Notably, to the extent there was a felt sense competition between OCA and Google, it was very one-sided. For all the OCA’s anti-Google rhetoric, Google never engaged in any parallel bashing of the OCA. In part, this was because it just didn’t need to – the OCA was much smaller, and never posed a threat to Google’s aims – and in part, Google also seemed to take the attitude of “the more the merrier” with regard to digitization (especially, one suspects, when the additional projects are willing to open all their data to Google’s search robots, as the OCA was).

reach, mostly for free, and who can provide the best set of procedures and structures for achieving that access. This is rather an extraordinary fact, if you think about it: with all the social, economic, and technological changes that have occurred between the turn of the twentieth century and the turn of the twenty-first, giving information away to the public has remained a viable basis for competition among entities and individuals in relatively powerful positions: city councils, technology companies, universities. The question of why exactly this should be such a powerful and long-lasting phenomenon, I suggest, would be an excellent starting point for further social or even philosophical analysis.

1.4.2 Financial Gain

The issue of certain actors in these cases being motivated by financial gain is simultaneously obvious and contrary to the popular imagination of how these initiatives have played out. Of all the players discussed here, after all, the one by far most frequently accused of entering this domain in order to reap financial rewards has been Google. And yet, based on the data presented in the Google Books case chapter, I have to conclude that profits were far below the top of the list of Google's motivations to start scanning books. If profits even figured at all, they seem to have done so in an extremely circuitous and indirect way: digital books (and their readers) would provide the company with more data that they could use to improve their algorithms, and through that would help to buoy their central profit center in advertising. But that chain of reasoning seems to have been barely at the fringes of the Google leadership's mind, if even there, in the early years. And in fact, once the project did shift into a more corporate, profit-focused gear a few years in, it immediately began to wind down: after all, there are far easier and less expensive ways for Google to serve its bottom line than by scanning millions of books.

No, the actors for whom financial gain was a compelling motivation were not at the center of any project, but at the edges; the ends of the spokes, as it were. On some level, the Carnegie Library program, the Google Books Library Project, and the Open Content Alliance were all networks, linked through central hubs: Carnegie, Google, and the Internet Archive, respectively. And at the edges of those networks there were communities and libraries.²⁵¹ And those communities and libraries saw something for themselves in the hub, financially speaking. In each of these three cases, the individual or entity at the center of the project offered a compelling financial incentive to participate: Carnegie would provide the funds for a building; Google would cover nearly all the costs of digitizing entire collections or parts thereof; the Internet Archive would provide a more efficient, cheaper scanning method that would save libraries considerable money over the cost of doing it themselves. In all of these cases, the communities or libraries could have gone another way – not sought

²⁵¹ The Boston Public Library is not discussed here because it does not share this basic network form with the other three cases; rather, it was and is the more unitary and unified project of a localized set of individuals.

or taken a Carnegie grant, not partnered with Google, not collaborated with the Internet Archive – and many peer entities (other communities, other libraries) took those alternative paths. But for those who did not, the cost coverage or cost reduction on offer from the center of the project was a significant factor in the decision to pursue the other motivations noted here – the grand ideals of education, equality, and uplift, the competition with peers – in these specific ways. Money can be a powerful motivator, especially for those who are relatively resource-poor – not a novel or surprising observation, perhaps, but an important one to note in the context of these initiatives, particularly since the public discussion surrounding the later cases has scarcely recognized this factor – or at least has scarcely recognized the loci of its actual influence: not on Google, but on its library partners, and on those involved in the OCA as well.



The balance of these three elements – personal passions, principles, and pragmatics – is different within each case. All had all three, but at varying levels of strength. The principled arguments come through especially strongly in the public library cases and in the OCA – and this makes perfect sense. For the early libraries, these principles had not yet been articulated – much less become part of the social fabric as they are today – so they needed to be stated, loudly, clearly, and frequently. And for the OCA, principles formed its key point of differentiation from the Google Books Library Project – thus, again, the principles needed to be clearly articulated so that they could form the nucleus of the initiative. Indeed, Kahle would eventually go so far as to state, in my interview with him, that the OCA was *only* “a set of meetings and a set of principles” (Kahle 2011) – and although others differed with that assessment, it is not difficult to see how the project and its principles could become so elided, considering the strength of the connection between them.

For Google Books, passions and pragmatics played stronger roles. On some level, Google Books happened because Larry Page was very interested in it happening, and he happens to be one of the wealthiest people in the world, at the helm of one of the world’s most advanced tech companies. But it also came to pass in no small part because a set of libraries recognized, pragmatically, that partnering with Google provided an opportunity to further their institutional missions to a far greater extent and at a much faster pace than they could ever hope to achieve on their own. For the libraries, achievement of mission was certainly a piece of the calculus, but the factors of scale, pace, and cost coverage loomed at least equally large. After all, as noted, there were many other ways in which these libraries could have pursued their missions – just perhaps not as quickly, as grandly, or as cheaply-to-them.

2. Definitions

If only one thing has been made clear by the foregoing chapters here, it is that even when initiatives strive to serve “everyone” or gather together “everything,” true universality in

either sense inevitably proves chimerical. Intentionally or unintentionally, boundaries on use become established – perhaps not in the sense of outright prohibitions or recruitment, but certainly in the sense that different sorts of users are made to feel differentially welcome, or are differentially able to make effective use of the resources on offer. And in most if not all cases, these inequalities among users can be linked to the intentional actions of project leadership, and to the assumptions that lay behind those actions, even where no bias or differentiation among users was the explicit aim. Indeed, because they are often unintentional by-products of actions directed toward other aims, the dividing lines among potential users are often invisible or difficult to discern from the perspectives of those erecting them – but such lines exist and impact usage nonetheless.

The universal book collection, moreover, has proven equally as elusive as the universal audience. The properties of books, whether physical or digital – their technological affordances, their physical locations, their condition, their age, and more – have effectively obstructed the goal of assembling any truly comprehensive, universal collection. Indeed, prior to the advent of digital technologies, few attempts at such comprehensiveness were ever made; their rarity is such that even thousands of years later, the Library of Alexandria still serves as the primary touchstone for those seeking to pursue this type of project. In the digital world, the goal of gathering together all the world’s books has become much more plausible: digital books can be compactly and redundantly stored, and digitization does not necessarily require the permanent removal of the volumes from the institutional or individual collections of which they are presently a part.²⁵² But as Google and the Internet Archive discovered, even in the digital age, gathering “all” the books remains an enormous challenge, and one that has thus far gone unmet by either project (or any other).

The questions asked about these issues of universality at the outset were as follows:

- RQ2: How did those responsible for casting the initial shape of each project conceive of what they were doing – or hoping to do? Put differently, what was each project’s native concept of itself? In particular:
- How many and what kinds of people did the project’s founders think would be using their creations?
 - How comprehensive was the collection intended to be, and how was it shaped by policies regarding what to include and/or exclude?

The answers to some pieces of these questions – the numerical sizes of the user bases and of the collections – are quite idiosyncratic by case, and will not be fruitful topics for synthesis here. And while the case chapters mainly focused on the positive answers to the remaining questions – what types of users were privileged; what types of collections were sought – the

²⁵² Whereas the Library of Alexandria was largely built on the wholesale confiscation of books from all ships entering the port (e.g., Erskine 1995, 39).

two sections that follow will briefly sketch the boundaries of those sets on a more holistic level: where were the edges of “everyone”? And what got left out of “everything”?

2.1 Boundaries on “Everyone”

As a starting point for illuminating these boundaries, it will be helpful to recall the user-related topics discussed in each of the case chapters, above. The topic headings for each appear in the table below:

Boston	Carnegie	GBLP	OCA
• Everyone...in Boston • Those in need...but only the deserving • Both casual readers & serious researchers • Practitioners of the useful arts • Women • Immigrants	• Aspiring working class • Children • Racial Divides • (Also some of the BPL users, noted in intro to section)	• Global public • The Consumer • The Scholar • The Genealogist • Not-so-favored users: those w/o connectivity/digital literacy	• Anti-Definition • The (Internet) Literate User • The Maker • The Machine

Several lines of differentiation among users can be drawn from the above – some applicable across all cases, others applicable to only one or two. In some instances, later cases seem to have learned the lessons of earlier ones, and altered their boundaries accordingly. And notably, the means through which differentiation occurred – the regulative forces enacting or enforcing inequalities among users – also differ between the public library cases and the digitization cases.

In both the Boston Public Library and the Carnegie Libraries, the main sources of constraint vis-à-vis users (what types of people were allowed and/or welcomed) had to do with regulations and norms. Local laws and library policies dictated who would be allowed into the library and who would be allowed to use it in which ways (i.e. in-person use vs. borrowing), while local social norms constrained the sorts of people who would feel welcome entering or using the institution. While structural concerns were not absent – for example, individuals with physical disabilities may have found themselves unable to reach the entrance, up a flight of stairs, or individuals who only read languages other than English may not have found any books available that they could read – the dividing lines among users were much more strongly drawn by policies and norms.

Within the digitization cases, by contrast, the constraints on user base are primarily structural: a website is provided, and users are either able to use it effectively or they are not – and unlike in public libraries, there is no staff present to ask for help. In these cases, norms and rules have become embedded within and hidden behind the interface; they become invisible until interrogated. Neither Google nor the Internet Archive is generally making explicit rules about who can or cannot use their interfaces, and yet the assumptions and decisions behind those interfaces nonetheless tend to privilege certain users and

exclude others. Indeed, the regulative power of digital structures in these cases, as well as the normative and political constraints of the historical ones, recalls Lessig's theory of things that regulate, as articulated in *Code* (and its revision, *Code 2.0*), in which he suggests that four primary forces regulate behavior – norms, laws (or formal rules), markets, and architecture – and argues that within the digital world, code (that is, digital architectures) may be the most powerful (Lessig 2006).

Additionally, it is also worth noting that for all involved in both GBLP and OCA, the primary target users, at least on a conscious level, were the *preexisting* users of the various institutions involved (Google, IA, the libraries). For Google, the consumer. For IA, the tinkerer. And for the libraries, by and large, the scholar and the student. For all of these institutions, but especially for the libraries, there was a sense that by striving to serve their core existing user base, while also throwing the door open to everyone else, the public would be served automatically, without any special or additional effort being made to target alternative demographics (e.g., P8). And indeed, the prevalence of genealogists as users of these digital collections – not a favored user group within academic libraries at all – would tend to provide some evidence that this approach has at least some merit.

Still, in all the cases, one senses just below the surface (or, in the case of Carnegie, directly upon the surface), an assumption that the users – the public – will either already be like the creators of the initiatives, or will *want to become* like the creators of the initiatives – or partly both. Within the public library projects, this assumption is fairly explicit: the social mobility aim of the early American public library was strongly tied up with paternalism, ideas about social Darwinism, and an elite view of what the poor and the working class should aspire to become. Within the LSDIs, these issues are much more implicit, but remain present. The leaders of these two digitization initiatives rarely say outright that they are imagining people like them or their families (although some actually do: e.g., P4, P10, P11). And yet, frequently, when describing the uses that people might make of the resources, they describe the uses that they themselves, or others like them, might wish to make – like Brewster Kahle wanting a content platform to seed new interfaces, or certain partner librarians envisioning a bonanza for in-depth research – and the interfaces provided seem quite likely to support those uses best.

Beyond these generalities, however, six more specific boundary lines can be identified among users of these four cases (though not all of the six boundaries apply to all four cases): age, class, race, gender, literacy/literacies, and the digital divide. These themes are each explicated in brief below.

2.1.1 Age

As discussed in the chapter on the Boston Public Library, one early dividing line among public library users lay between adults and children. The BPL, at its initiation, was

envisioned as an institution that would fill in the gap left after individuals aged out of the public school system; a place where *adults* could go to continue learning. As such, when the first dedicated building for the library was constructed, it contained no dedicated space for children, and no collections specifically targeted at a youth audience. The children, after all, had the public schools to provide for their learning; the library was for adults. And yet, this did not prevent children from attempting to use the library – far from it. But those who did so not only found themselves inadequately served, but also made the library more difficult to use (noisier, more crowded, etc.) for the adults who were the target audience.

The lesson that public libraries would attract children, and that this could even be a central benefit of such institutions, was quickly learned. The second BPL building constructed, in Copley Square, included a dedicated children's room, and by the time of the Carnegie library program such spaces had become *de rigeur*, and were included in all but the two smallest library plans suggested by Bertram to recipient communities. Whether this lesson has been taken to heart by the LSDIs examined here is far from clear. Certainly, there is nothing explicitly barring a motivated child from utilizing the digital books made available through either project – and yet, given the collections being scanned, one wonders how much age-appropriate material such a child is likely to find. Academic and research libraries, after all, are generally geared exclusively toward adults – college-age and above. Some children's books and young adult literature undoubtedly exist in these collections, but the proportion is small. Still, if the public library movement offers a lesson in this area, these and future digitization projects may be well-advised to take seriously the needs of young users, or risk underserving what could be a core demographic going forward.²⁵³

2.1.2 Class

The issues of class in the public library movement have been well discussed in earlier chapters, and will not be dwelt upon here. However, it does bear briefly restating the fact that although both the Boston Public Library and the Carnegie Libraries were explicitly – or at least rhetorically – intended to uplift the lower and working classes, and to increase class mobility, there is considerable evidence that in practice, working class users were often made to feel unwelcome by library staff, in ways that significantly undercut the libraries' stated rhetoric. Indeed, Harris (1975a) identifies this condescension as one of the major failings of the public library movement, and both Garrison and Macleod have devoted significant portions of book-length works to explicating how these exclusionary practices played out on the ground (Macleod 1968, Garrison 1979).

The paternalistic rhetoric of class uplift is far weaker in the digitization cases, but still comes through in some of the partners' statements about their responsibility to “share the

²⁵³ As, indeed, some other digital libraries have – for example, the International Children's Digital Library: <http://en.childrenslibrary.org/>

wealth:" if such a responsibility exists, there must be "have-nots" with whom that wealth needs to be shared. In the digitization cases, that would be anyone without physical access to or permission to utilize the libraries whose collections were being scanned – exclusions which are not identical with, but often run in parallel to, current class divisions. And there is also a more explicit class divide among the user base of these initiatives – but it takes a somewhat different form, and will be discussed in the section below, on the digital divide.

2.1.3 Race

The only case of the four examined here in which race emerges as an explicit factor impacting use is the Carnegie Library Program. In that program, as previously explained, Carnegie's requirement that building requests come through elected local officials formed an effective, if unintended, barrier to the provision of Carnegie libraries to the black population of the American South. Essentially, because all elected officials in the U.S. at the time were white men, and the white population of the South was not especially sympathetic to the needs of its black neighbors, it was much more difficult for communities to obtain so-called "Negro branches" than it might have been if Carnegie's policies had allowed requests to flow through some alternate channels, beyond elected officials.

Within the other three cases, there is virtually no discussion of race whatsoever, and any speculation about the role of race in those cases would be just that – speculation. But that does not mean it was not a factor. In the early years of the Boston Public Library, it seems plausible to guess that the lack of discussion of race as an issue had much to do with the relative racial homogeneity of the local population – although there are many accounts of an expanding immigrant population, most of these immigrants were coming from Europe. Under such conditions, it is perhaps unsurprising that racial differences *per se* did not come up in official documentation of the library's early years (though immigration issues did): white privilege in the U.S. is such that those who have it simply don't need to think about race on a regular basis, and so they do not do so – and they thus also tend not to commit thoughts about race to writing in official city reports. All of this is not to say that race was not an issue in Boston – it may well have been – but it was an issue invisible to the eye of the historical researcher, given what has been documented and preserved.

Within the digitization cases, on the other hand, it seems more than likely that there were no intended racial divisions in the user base on the part of anyone involved, and the administrative processes and technological structures of the initiatives do not appear, at least to my eye, to privilege or disenfranchise any particular group specifically on the basis of their race – at least no more than any of the institutions participating in the projects do, simply by virtue of existing within a time and set of places where social inequalities based on race are still ubiquitous. Racial boundaries have grown much subtler and less explicit from the nineteenth century to the present, but that hardly means they have disappeared,

or that they could not be working some subtle differentiation in usage practices surrounding LSDIs.

2.1.4 Gender

The issue of gender also arises mainly within the Carnegie Library program, and in a way very similar to the issue of race. Women could not vote or hold office either, and so Carnegie's policies mandated that their voices be mediated through those who could – i.e. men. Thus, at the point of making requests to Carnegie for donations, a great deal of female effort and agitation became co-opted by local male leadership.

However, that co-option at the point of request is only one part of the story, and not the part most relevant to *usage* in any of the four cases. Rather, in all four cases, women have always been quite welcome as users, on a par with men. Given the time period of the two public library cases, in fact, the level of gender equality with which access was offered was quite remarkable. And of course women also comprised a large percentage of the front-line staff of these institutions as well, and even served as high-level leadership in many Carnegie-donated institutions ("Free Public Library: Atlanta, Georgia" 1898, correspondence between James Bertram and Tommie Dora Barker, Garrison 1979, 173, Jones 1997, 42-43).

Interestingly, although as with race, the digitization projects appear to exhibit no greater gender discrimination among their users than exists in the contexts of use already, the administration of these projects to date is quite a bit more male-dominated than public libraries have been since at least the turn of the twentieth century. This issue could easily be the topic of another dissertation; for the purposes of this one I will simply note that while librarianship quickly became a stereotypically female profession, computer programming has equally quickly become a stereotypically male one – and the digitization projects have tended to reflect a higher proportion of computer programmers' visions (Larry Page, Brewster Kahle...) than librarians'.

2.1.5 Literacy/Literacies

The issue of literacy – or more accurately, *literacies* – acts as a boundary on use for all of these cases. On a very basic level, effective use of a large collection of books requires that one be able to read (and specifically must be able to read in the language(s) in which the books are written). Public libraries, of course, have proven excellent instruments for the teaching of reading skills, especially after their early turn to serving children, discussed above. And perhaps LSDIs can begin to serve this purpose as well – although one suspects that the generally high-level academic bent of their existing collections would tend to diminish their utility for the teaching of basic literacy. Overall, although it may not always be necessary to be reading-literate in order to begin to use a library or an LSDI collection, users possessing such literacy are undeniably privileged in both contexts – as indeed they

would be in any context where the central aim is to provide individuals with thousands or millions of mainly text-based volumes.

And yet, this basic form of literacy is far from the only one required in order to use either a library or an LSDI. Rather, each of these types of institution or project surround the textual materials they provide with all manner of intellectual scaffolding which must also be negotiated in order to obtain effective access. Human beings, after all, are no more born knowing how to use a card catalog, a classification system, a search engine, or a wiki than they are born knowing how to read. All of these things must be learned – and in the types of project examined here, these types of technological literacy effectively form a moat (or a wall, or a fence – pick your metaphor) around the actual texts to which traditional reading-literacy can be applied.

Consider the library catalog as an example. In mid-nineteenth century Boston, the BPL was the first institution to introduce the technology of the catalog to the public – first in bound form, and then in cards. By the late twentieth century, training in the use of such systems had become a standard part of elementary education, but when the BPL first opened, that standard had not yet been enacted. Pair this with the fact that offering guidance to readers was not even mentioned by library leadership in the extant historical documentation until more than ten years after the library first opened, and one can begin to imagine how perplexing these catalogs might have been for the average member of the public. In theory, and technically, the books held by the BPL were free and open to public use; and yet, in order to obtain actual and effective access to those books and their contents, users had to be not only able to read, but also able to navigate the technology of the catalog (and in the case of the card catalog, this would be quite a cutting-edge technology indeed). It is difficult to imagine that this would not have presented a barrier to many potential users.

Within the LSDIs, the same story plays, though the technology has changed. From the relatively elite perspective of technology companies and major university and research libraries, it can be difficult to keep in mind that the web utilities we live and breathe on a daily basis – browsers, WiFi, search engines, and more – are not actually things one is born knowing how to use. At some point, everyone has to learn, and far from everyone has done so. Literacy in these technologies plays the same gatekeeping role as did catalog-literacy in the early public library. Yet, this gate is only one piece of a much larger barrier – the digital divide – which is the last boundary on users that will be discussed, below.

2.1.6 Digital divide

The concept of the digital divide combines the issue of inequities in technological literacy as described above with the issue of inequities in functional, physical access to technology. The essential point is that (a) not everyone has access to a computer and/or the internet, and (b) even among those who do have such physical access, not everyone is able to make

effective use of those tools. Problem (a) is especially acute if one looks beyond the United States. According to the Pew Research Center, as of 2013, 85% of U.S. adults were using the internet (Zickuhr 2013, 2). But worldwide, the International Telecommunications Union (ITU) places the figure at only 39%, with significant inequalities by region, national economic development, and gender (International Telecommunications Union 2013, 2). A summary of the ITU statistics for 2013 appears in the table below (International Telecommunications Union 2013).²⁵⁴

	% Using Internet
Global Population	39%
Men	41%
Women	37%
Economic Development Level	
Developed World	77%
Developing World	31%
Economic Development Level + Gender	
Developed World - Men	80%
Developed World - Women	74%
Developing World - Men	33%
Developing World - Women	29%
Region	
Europe	75%
The Americas	61%
Commonwealth of Independent States (former Soviet)	52%
Arab States	38%
Asia & Pacific	32%
Africa	16%

Notably, within these statistics, we can observe several of the divides listed above – especially class, race, and gender – being projected into a new context, on a global scale. In the developed world, the vast majority of individuals – 77% – have access to the internet; in the developing world, by contrast, a similar majority – 69% – do *not* have such access. Further, while a gender divide in access exists worldwide, that division is much starker in terms of actual population counts in the developing world vs. the developed world – as the ITU notes, 16% fewer women than men use the internet in the developing world, and that percentage corresponds to 200 million people without access. And although obviously the issues of race on a global level are much more complex than in the U.S. alone, on a very basic level the breakdown of access levels by geographic region indicate that there are significant racial inequities in internet access as well – though these of course intersect with

²⁵⁴ All terms and definitions theirs. The ITU takes their definitions of developed vs. developing countries from the United Nations, as noted here: <http://www.itu.int/en/ITU-D/Statistics/Pages/definitions/regions.aspx>

issues of nationality, citizenship, immigration, and diaspora, and cannot be disentangled within this data as presented.

In the context of LSDIs, this is a very significant barrier to usage – if technological literacy is a fence or a moat, the divide in actual, physical access to the internet can look more like the Berlin Wall. Put quite simply, if you are not on the internet, you are not downloading content from Google or the Internet Archive – whether that content is a book or any other piece of information or media. And if part of the purpose of LSDIs is theoretically to help equalize access to information worldwide, these inequities in internet access present a colossal barrier to achieving that goal: after all, they prevent precisely those likely to be most information-poor from accessing the high-quality information provided through LSDIs.

Now, of course, the digital divide was not created by those promoting digitization, and it is not necessarily their responsibility to fix it. And it can quite reasonably be argued that even if the output of LSDIs cannot reach 80% of the population in some parts of the world, well, those individuals are at least no worse off than they were before, and might still be better off. After all, a plane ticket to one of the institutions doing the scanning is unlikely to ever be affordable by most of the world's population, whereas internet penetration is steadily increasing over time, and plausibly could one day reach a much higher percentage of the world. To put it in more human terms, the motivated student in sub-Saharan Africa may not currently have access to Google Books, but she might someday, and maybe even soon. And of course, that scenario is much more plausible than that same student ever walking through the doors of Stanford's Green Library building.

2.2 *Boundaries on "Everything"*

In any collection of books, there is a tension between selectivity and comprehensiveness – and both of these factors depend heavily on the resources available for building the collection. More money, and one can gather lots of books, and spend time making sure at least most of them are *good* books. Less money, and the collection grows smaller, and the beggars-can't-be-choosers maxim begins to apply, to the detriment of collection quality. Of the four cases examined here, it seems fair to submit that the Boston Public Library exhibited the highest level of selectivity – both specifically seeking to provide high-quality materials and working within limitations based on physical space – while the Open Content Alliance exhibited the lowest: the Internet Archive has always been willing to scan and/or host just about any content brought to them. And there can be little question that the Google Books Library Project is the most comprehensive of the four, with its 20+ million volumes and counting, while it seems equally clear that the smaller, retail-period Carnegie Libraries were the least comprehensive, often scrambling to collect whatever books could be found in local attics just to have something to put on their shelves.

In all of the cases examined here, the major constraints on collection scope have had much more to do with pragmatics – what the collections *could* contain – than with any kind of ideals or selectivity about what the collections *should* contain. And yet, the relevant pragmatic considerations were significantly different between the public library cases and the digitization cases (though some overlap does exist).

Within the public library cases, the technology of the physical book forms the primary constraint on collection scope. Most saliently, a physical book can only be in one place at one time. In order to provide the same book to many people, one needs many copies of the book, or a very patient group of people (as they would each have to wait to use the same copy in sequence). In very basic terms, the technology of the paper book means that physical libraries have always had to choose between building a consolidated, massive collection in one place or building smaller, redundant collections in many places. Urban centers with a central library and a network of branches hybridize the two to some extent, but still over a relatively localized area. At a basic level, Boston took the approach of building a large, consolidated collection (and only later a local branch network), while the Carnegie Library program took the approach of seeding smaller collections throughout the English-speaking world.

A key reason why book digitization is so radically transformative, then, is that it pushes the book past its physical limitations. When a book becomes a chunk of binary data, and a global data-transmission network exists to re-copy and transmit that chunk of data anywhere and everywhere, that book is no longer fixed to a single time and space. This allows the best of the options available to physical libraries to coexist in a way not possible in the physical world: a massive number of digital books can be offered to a massive number of people over digital networks, all at the same time. Still, having removed most of the constraints presented by the physicality of the paper book, others emerge – particularly at the point where the conversion from print to digital is taking place. At that conversion point, the physicality of the book creeps back in, along with the at-least-equally-significant constraint of copyright law.

The physicality of the books matters to this conversion process in at least three ways: location, format, and condition. Before books are made digital, they are still strongly tied to a specific location – and that location must be identified in order to find and scan the book. One of the major constraints on comprehensiveness for both Google Books and the Open Content Alliance – though Google Books has been more strategic about trying to get past it – was the fact that books exist in thousands (maybe even millions) of pockets worldwide, some large and some small. In order to scan every book in the world, one must first *find* every book in the world. And that is an extremely non-trivial venture. Google went about it by first signing on a group of very large libraries and then trying to fill in around the edges with unique collections; that seems as good a strategy as any. The Open Content Alliance,

and later the Internet Archive, never took such a systematic or center-oriented view; it simply scanned whatever came in – thus increasing its collection’s levels of idiosyncrasy and duplication, and reducing its likely comprehensiveness.

The format and condition of physical books also have an impact on whether they can be made digital. Both Google’s and the Internet Archive’s fast-paced scanning machines have limitations regarding the size of the volumes they can accept, which in turn limits the content of both collections in fairly unpredictable ways, as books larger or smaller than the scanners can handle are excluded. The Internet Archive, it should be noted, is willing to scan those large or smaller books as well, but using slower-paced machines and at additional cost to the source library; this likely diminishes the gap introduced by book size, but does not eliminate it completely. And finally, books in poor condition – very old books, books printed on acidic paper, books that have not been properly stored, books that have been extremely heavily used – also often get left out of mass scanning projects, because either the source library or the institution doing the scanning balks at the possibility of further damaging or even destroying such books by putting them through the scanning process.

Even once books have been converted to digital form, however, there is still one extremely important constraint placed on building – and even moreso, on sharing – the collections. That is, copyright law. It is in this one area that digital books become more complicated and often less conducive to sharing than physical ones. A physical book, after all, does not need to be copied in order to be shared: one just hands the book over to the person wanting to read it. A digital book, however, is different. The scanning itself makes a copy. The transmission of the scan to a server for hosting makes another copy. The transmission of the scan over the Internet to a user’s screen makes yet another copy – and so on and so on. Now, for public domain materials, all of this copying is unproblematic; it is only for materials in copyright (generally speaking, newer materials not published by the government or otherwise rights-free) that it becomes an issue.

That is, although it is *technologically* possible for LSDIs to provide the full text of millions of books to millions of people all at once, it is not *legally* possible to do so except where permission has been granted or where the works are out of copyright. And so, from the user perspective, despite all the noise made about Google’s disrespect for the copyright system throughout this project, the *usable* portion of the collection digitized from libraries – the part users can actually *read* – does not include works in copyright. Those works are on Google’s servers, but they cannot be shared; for Google to do so would be a clear violation of copyright law (whereas the scanning alone has been deemed a fair use (Chin 2013)). The Internet Archive does provide some access to in-copyright works for reading, through its lending program, but those services too are carefully circumscribed to analogize to physical contexts (ironically, reinstituting the limitations of the physical book – one copy to one

reader at one time) so as not to further raise the ire of copyright holders and their advocates.

To summarize, the collection and access scope of each of the cases presented here can be thought of as follows:

Case	Scope
Boston Public Library	Large volume of books to a relatively small number of people.
Carnegie Library Program	Small volume(s) of books to a relatively large number of people.
Google Books Library Project	Huge volume of books to huge number of people – but free access only to very old materials.
Open Content Alliance	Large volume of books to large number of people; free access generally only to very old materials, except through various e-lending schemes.

3. Implementation: Processes and Structures for Access at Scale

At the point of implementation, the motivations and definitions behind an information access initiative (or any other kind of sociotechnical endeavor) coalesce, moving from the theoretical realm into the material reality of actual practices and structures. Each of these practices and structures reflect intentional processes on the part of some actor – or, more typically, several actors – but each one also escapes the full control of those actors once enacted (in the case of procedures) or built (in the case of structures). At the outset, the questions asked regarding this piece of the information democratization efforts examined here were as follows:

RQ3: How did each project initially go about achieving the goals and definitions outlined in RQ2, particularly in their implementations of particular processes and structures? In particular:

- What procedures were put in place to help achieve the goals outlined in RQ1 and definitions outlined in RQ2?
- What design moves were made to orient the structures toward their intended user base and level of comprehensiveness?

To an even greater extent than the responses to the definitional questions posed above, the responses to these questions of implementation reveal an enormous amount of idiosyncrasy: each of these projects unfolded in its own unique way, following its own very specific and situated set of practices, and erecting its own tailor-made interfaces and backend structures. These idiosyncrasies have been discussed in great detail in earlier chapters, and will not be rehashed here. Rather, the sections that follow will focus on extracting common threads across the four cases; procedures and structural elements that seem to transcend the socio-historical context of the initiatives, and begin to appear necessary, or at least typical, within projects that share these initiatives' massive, idealistic goals regarding the expansion of information access to the public.

Because the processes and structures of implementation are so firmly interwoven in each case, moreover, the analysis that follows will treat both together. As a jumping-off point for doing so, the table below presents the procedural and structural subheadings used in each of the case chapters:

Boston	Carnegie	GBLP	OCA
Processes			
• Management (incl Leadership Politics, Committees) • Acquisitions • Cataloging • Public Service • Protecting the Collections	• Early paternalism • Corporatization ◦ Formalization: conditions and paperwork ◦ Authorization: Public support and who can declare it	• Selection & Ordering • Library pre-processing • Transit & Scanning • Digital Delivery • Constraints on use	• Admin Structure (or lack thereof) • Scanning • Constraints on End-use • Transparency in practice
Structures			
• Buildings • Stacks • Catalogues	• Building planning ◦ Branch libraries • Stacks arrangement	• Google Interface/Views • Library hosting/serving • Standards (image/metadata) • Invisible structures	• IA/Open Library Interfaces • Platform Approach • Standards (image/metadata)

Among the widely varied processes and structures implemented in each case, four major threads come through as especially crucial to the success of this type of project. The first three will be treated roughly in order from most procedural to most structural, with the fourth representing more of an absence, and also a direction for future work. These issues are as follows:

1. **Leadership.** A particular leadership trajectory occurred in parallel in three of the four cases: an initial spark kindled by visionaries with little practical idea what they were getting into, which was then blown into flame by on-the-ground leaders to whom the pragmatics of each project were delegated. Notably, the fourth case, the OCA, falls out of parallel at the point of delegation, and largely as a result of this, never fully moves past the “spark” stage.
2. **Standardization and Systematization.** This factor came into play after the point of ignition, when each initiative was striving to rapidly increase in size. At that point, each project began to routinize and standardize both its processes and its structural elements, to make it possible to perform the tasks that needed performing in greater bulk and at greater speed.
3. **Structural Residue.** Of these four projects, two have ceased to exist, and a third is winding down fast. And yet, even after these projects end, they leave behind durable traces of their existence that have meaning for our relationship with information resources well beyond the active timespan of the projects themselves.

4. **The Guidance Gap.** In the early years of all four cases, a significant absence exists in the area of user guidance: that is, in each of these cases, the leadership initially fixated on the pure provision of physical or technological access to materials, while neglecting to make any provisions for the kinds of assistance that users inevitably require when faced with such mountains of information resources. This is a significant issue particularly for the LSDI cases, and will be a fruitful ground for further research.

Each of these four themes are discussed in turn below.

3.1 Leadership

The discussion of motivations above introduced the concept of the personal passion spark: the observation that in each of the cases examined here, the personal history and predilections of at least one very wealthy individual formed a necessary, though not sufficient, condition for their initiation. The discussion of leadership processes will build upon that observation, suggesting that in addition to passion and money, a healthy dose of naïveté about the functional realities of providing lots of books to lots of people helped each of these projects achieve liftoff. More specifically, as I have argued elsewhere (Jones 2013b), whether or not naïveté is *necessary* to the initiation of large-scale information access efforts, it can certainly be *helpful* – and yet beyond the point of initiation, the ultimate success of such projects depends heavily on the empowered involvement of individuals with significant expertise and/or ingenuity in working through all of the many details.

In each of the historical public library cases, the individual whose donations breathed life into each project – Bates for Boston, Carnegie for the Carnegie Libraries – exhibited little knowledge of the pragmatics of running a library, and little inclination to learn them. Both had simply concluded from their experiences as users of private book collections that public libraries would be valuable, and both left all the details up to others. For the BPL, some of these details were worked out by the library's Trustees, and many others by the library's early superintendents, C. C. Jewett and Justin Winsor, and other early staff. For the Carnegie program, the details of the donation process were largely left up to James Bertram and then the Carnegie Corporation board, while again, the on-the-ground running of the libraries – the actual procedural details of their implementation(s) – was organized by the librarians and other staff at those institutions. At this point, we can see how the mediation and power structures described earlier and the leadership processes outlined here collide – in fact, for the historical cases, they are virtually identical. To extend a common musical metaphor, the man who pays the piper calls the tune – but it is the piper himself who does the actual playing, and who needs to have the requisite knowledge of his art to do so. One need not know how to play to desire the music, and equally, neither Bates nor Carnegie needed to know how to run a library to wish (and fund) them into existence. And in these cases, the pipers are the staff actually present in the libraries, making things run, while we

could imagine the BPL Trustees and James Bertram as conductors, making sure the pipers' work flows smoothly.

Within the Google Books case, another layer becomes evident in this interplay between naïveté and delegation. That is, the relative early naïveté of Larry Page and other high-level Googlers regarding library practices allowed them to embark on a project that had repeatedly been discussed and dismissed by those in the library community – those with enough practical knowledge to “know” that digitizing all the books in the world was “impossible” and was therefore not worth trying. Indeed, three of the librarians interviewed for this case would describe in strikingly similar terms how, in the libraries' pre-announcement negotiations with Google, librarians brought forward myriad concerns from prior experience with book digitization, and the Google representatives essentially waved them off. For example, one recalled,

I'd come in and we knew some things about duplication, how many copies of Nature we had, how many copies between the branch libraries in the campus, what percentage of the collection was duplicated. So I start to talk about that. ... And Larry Page went “yeah, we don't care about that.” Because it would be harder to sort that out than just do ‘em. And we went [meekly] “OK!” (P9)

As experts within the existing paradigm of digitization, it was difficult for these library leaders to see beyond all the potential problems; standing outside of that paradigm, the Google teams (perhaps hubristically), did not share their hesitation. And although many of the librarians' concerns have certainly proven valid (especially with regard to copyright and metadata (e.g., Duguid 2007, Samuelson 2010)), as one of those librarians noted, Google's naïveté in these areas – and some librarians' willingness to play along with it – helped make the project seem less impossible:

These Google people were like these adolescents with these great ideas. And we were like the old folks, in a way. [But we decided,] “this is going to be so cool, let's let them do this...and when in doubt, let's do what they want to do. Let's think like they think. Let's not think the way we think, and have to have everything has to be tight and neat and you've got to make sure you've got the holdings record right” (P17)

By setting aside their strong sense from within the paradigm of librarianship that such scale was impossible – and specifically impossibly complex, impossibly expensive, and impossibly risky – Google's library partners were able to take a significant leap. And, it should be noted, just because the librarians were willing to go along with Google's model does not mean that they lacked input: in fact, as previously discussed, Google relied critically on the expertise of partner librarians, especially at Michigan, Stanford, and Harvard, to help them work through the procedural details of scanning tens of millions of books.

The Open Content Alliance, in this context, is the exception that proves the rule. As noted at many points, the driving personality behind the OCA was Brewster Kahle. And although Kahle did have a foot inside the existing library paradigm, having been scanning for years already, there is still a sense from interviews that coming out of the software industry, Kahle still retained some degree of outsider perspective. In particular, he felt that libraries moved too slowly, and although he believed the OCA could make things go faster, he had no clear plan for how the organization would independently function. As one partner librarian suggested, “he was just ‘Come on! Let’s just go scan some stuff!’ and there wasn’t really much thought of contract or MOUs, or you know, the usual stuff that define people’s relationships” (P3). Placing a more positive spin on Kahle’s approach, one IA employee observed that the organization was

in a liberated position...because we’re not beholden to the normal baggage that is associated with library systems and library practice, and none of us are librarians. We’re software people. You know? And that allows us to sort of approach the issue in a totally fresh way (P7).

And indeed, the IA has made solid strides in book scanning, as already discussed: its Text Archive is second only to Google Books in the size of its corpus, and the IA continues to scan (and otherwise accumulate) books at a fair clip. Still, the Open Content Alliance *per se* no longer exists. And interview accounts suggest that to a significant degree, it broke down because of Kahle’s unwillingness to delegate and share control. As one early participant recalled,

Me and a few others, really early on, had a discussion with Brewster about ‘look, you can run this as part of your Internet Archive, and it’s an extension of your ego, and that’s fine... or you can establish it as a kind of a trust, which has its own independent governance... and that’s fine too. But you can’t do both’ (P3).

And another concurred, adding that at a certain point,

[Our] understanding was that [Maura Marx] would start a new organization that would be an independent organization and would hammer out how to work together with the Internet Archive to, you know, to transition this thing. And... when it came down to it, though, . . . I think Brewster just wasn’t really quite able to let it go (P11).

At a certain point, some delegation and leadership-sharing was necessary to ensure the survival of the OCA, but it does not seem as though Kahle recognized the fundamental importance of that form of cooperation, at least with regard to his lifelong passion for building the universal library – and between that and the loss of Microsoft’s funding, the OCA collapsed. Kahle and the Internet Archive’s library-outsider enthusiasm may have been sufficient to launch the OCA, but they proved incapable of independently keeping it aloft, at least in its initial form.

Though each of these cases include unique manifestations of library-outsider naïveté, they are nonetheless similar in that in each case, at least one influential individual believed in the possibility of something unprecedented – a major urban public library, a constellation of library buildings across the globe, the world’s premiere research libraries made digital within a decade, open collaboration on distributed book scanning – and used their considerable financial and/or technical resources to help usher that unprecedented thing into existence. Still, the extent to which each of these projects has succeeded – or not – has depended crucially on the involvement of less naïve individuals in working out the actual specifics. Being innocent of existing library “truths,” it seems, can be a significant aid in dreaming big in the book-access arena; however, beyond a certain point, more detail-oriented expertise and ingenuity is required to transform those ambitious dreams into a new reality.

3.2 *Standardization and Systematization*

One of the key elements that sets these cases apart from other public library initiatives and other digitization projects is the sheer grandiosity of scale sought by each one. The Boston Public Library was the largest public library of its time, by far – and indeed, it remains one of the largest public libraries in the world even now. Andrew Carnegie sought not to donate just one library to his hometown, or a handful of libraries to places he liked, but both of those things *plus* thousands of other, smaller libraries, available to any community willing and able to put together a request meeting a few basic requirements, regardless of any personal connection to Carnegie himself. And neither Google Books nor the Open Content Alliance were content to scan just a few delimited collections as most digitization projects had before them; instead, they set their sights on scanning as many books as they could possibly get their hands on, almost regardless of the actual contents of those books. In order to meet or even approach any of these grandiose ambitions, all of these initiatives moved quickly from their *ad hoc* beginnings to adopt a more systematized and standardized set of procedures and structures capable of scaling up to meet the vision.

Such standardization has two core advantages for initiatives seeking scale: first, it facilitates delegation to others – and even large numbers of others – which in turn diminishes the impact of losing any one individual worker (in much the model of the assembly line, though two of these cases predated Henry Ford by decades). And second, it expedites extensibility: whether the project is growing through accretion of library partners, recipient communities, collection contents, or users, the development of standardized systems makes the requisite onboarding processes flow more quickly and smoothly. This process of standardization, however, manifested differently in each case.

For the Boston Public Library, the main processes in need of standardization were those required to expand the collection in different periods of its history. Early on, following the

second Bates donation, which funded the acquisition of nearly 25,000 volumes, then-Superintendent C. C. Jewett and his staff developed a standard workflow for processing and cataloging the materials (as illustrated in Chapter 3, Figure 24). This standard workflow not only facilitated the efficient incorporation of a donation that nearly doubled the size of the library's collection in a single year before it even had a building in which to house them, but provided the basic template for how acquisition workflows would function within the BPL for decades to come.

A second, and more structural, example of standardization from the Boston Public Library came later, in the transition from bound catalogues to card catalogues led by Justin Winsor in the early 1870s (Whitehill 1956, 97). Bound catalogues, as previously described, are very difficult to keep current as a library's collection expands – and especially as it expands beyond the capacity of one or two volumes. Card catalogues, by contrast, provide a highly extensible, scalable system for keeping track of even very large collections: as the collection grows, one can simply slip more cards into the proper locations, expanding the card set into additional drawers and cabinets as needed, without having to revise or reprint all of the previously-entered cards. Indeed, even despite the all-but-ubiquitous use of online catalogs in the present day, several card catalogs still remain in active use in some BPL departments.

Within the Carnegie Library Program, two major loci of standardization had to do with donation policies and library architecture. Both of these trends have been detailed at length in Chapter 4; as such, they will be only briefly summarized here. The donation policies began to become progressively more standardized early in the shift from the retail period to the wholesale period – that is, the major period of scale-up within the program as a whole. First, around 1897, the requirements that communities provide a plot of land and commit to tax themselves in the amount of at least 10% of the donation annually to maintain the library were introduced (Koch 1917, 11-12, Bobinski 1969, 187-88). Not coincidentally, the imposition of these standardized conditions coincided with the near-complete delegation of the program's management from Carnegie to Bertram. And then, with Bertram at the helm, the process of standardization – and arguably, bureaucratization – continued. A succession of formal request forms were developed, each one requesting more data from the communities than the preceding version. And at this point, the restrictions on who would be authorized to submit building requests, to elected officials only, were also introduced. Within a few years, the process had become almost totally mechanical – if a community could meet the (successively more stringent) requirements, they would be awarded a Carnegie grant to build their library. And in fact, it is difficult to imagine how the program could have made as many donations as it did – more than 2100 – with a staff of essentially only one person, without all these standardized processes in place.

In parallel to these procedural innovations, moreover, the Carnegie program also began to advocate for changes in the architecture of the types of small library buildings that were

generally erected using Carnegie donations, culminating in the articulation and dissemination of the six standard building plans discussed earlier. Although the use of these plans was not required by Carnegie or the CCNY, many communities used them nonetheless, leading to the erection of hundreds of very similar small local and branch libraries in the later years of the program. In this piece of standardization, one can perceive two major gains in efficiency: on the one side, these plans were often a more efficient use of funds than the library plans developed independently within small communities, where there was little expertise in this type of design to be found; this, of course, was the efficiency gain that mattered most to Bertram in disseminating the plans in the first place. However, on the other side, within the communities, there was an additional gain in efficiency to be had from these plans: that is, if they simply used one of Bertram's plans, they could cut through much of the red tape on getting their Carnegie donation approved, and simultaneously circumvent the often drawn-out process of soliciting and then choosing between local designs (as occurred repeatedly, for example, in Boston). Essentially, it allowed these communities to leapfrog to a more efficient design, developed out of lessons learned from less successful earlier building donations (e.g., Van Slyck 1995, 146).

Within the Google Books Library Project, the library partnership contracts and the scanning process formed the two major loci of standardization. And although the Open Content Alliance generally abhorred contracts, it also progressively developed more standardized scanning and materials ingest procedures. Thus, while the issue of contracts will be discussed solely in relation to Google, the discussion of scanning procedures will incorporate elements from both GBLP and OCA.

The first five Google library partner contracts were negotiated out of public view, in isolation both from one another and from other libraries and the public. Four out of five of them, in fact – all but Michigan's – have remained hidden. Still, from the interviews conducted for this dissertation as well as other statements and observed actions taken by these five libraries, it is clear that each of these contracts was fairly unique, and each seems to have included more concessions to the library's desires and perspectives than later Google contracts would. However, negotiating contracts this way – one at a time, with negotiating power on both sides – can be very time-consuming. As a point of reference, roughly two years elapsed between Larry Page's first discussions with the University of Michigan Libraries and the signing of that library's actual Google scanning contract (P8, P10, University of Michigan and Google 2004, "MDP Timeline") – and two years is a veritable eternity in internet time. As such, it is unsurprising that beyond those first five libraries, as the project scaled up, most of the other library contract negotiations followed a very different form: essentially, Google offered a standard contract, and aside from minor modifications, libraries could take it or leave it.

There are two observations worth making about this shift. First, it makes sense in the context of the project overall that libraries beyond the first five would have less power. The first five, after all, were necessary to have on board not only to get the project started, but to kick it off with the desired brand identity: these were not just any libraries after all – they were five of the most elite library collections in the world, and they said yes to Google. That endorsement was extremely powerful and useful to Google at that early point. But beyond that point, the partnerships became a buyer's market for Google, where the company could target institutions based on specific collection objectives and then largely dictate the terms of the contracts, to take or leave. And that brings us back around to the second observation: that is, the standard contract allowed for *much* faster turnaround on library negotiations by Google. Where the first five contracts took more than two years to completely negotiate (2002-04), the next three years of the project (2005-07) would see the addition of *twenty-four more* library partners – a radical expansion in scope.

The final noteworthy element of standardization within both the Google Books Library Project and the Open Content Alliance, as noted, has to do with the actual processes for scanning the books. In both of these cases, one of the very first steps was the development of new scanning technologies – Google's multi-camera, algorithm-oriented system on the one hand, and the Internet Archive's Scribe scanner on the other. Both of these systems allowed books fitting their specifications (not too big, not too small, not too fragile) to be scanned at a very rapid pace and a very low cost (partially due to the rapid pace). There is no question that without technologies like these,²⁵⁵ neither project would have been at all feasible. And beyond this, both projects also had to work out systems for physically bringing together the books and the scanners, and those systems had to run as quickly and efficiently as possible. For Google, this meant a network of regional scanning centers, fed from libraries by fleets of trucks filled with special proprietary book trucks, which themselves had been loaded with books already cleaned and barcoded by the staff at their libraries of origin. For the Internet Archive, where libraries had contracted with it to perform the scanning, the procedure was quite similar, except that the scanning centers were often inside the libraries and open for examination by library staff and others.²⁵⁶ Coordinating these logistical issues was at least as important as working out the scanning technology – indeed, Google even hired a logistician to make it all work. Even a completely perfect scanning system, after all, would be incapable of scanning millions of books without an organized, streamlined procedure to keep the flow of books to the scanner in motion.

²⁵⁵ Clearly, they needn't have been *these specific technologies* – there are many other scanning technologies out there, and some are quite comparable in speed and efficiency to the systems used at both Google and IA. At the same time, most of these technologies are proprietary, and extremely expensive. Given the scale at which both projects were intending to operate, it thus made good sense to independently innovate rather than investing heavily in solutions produced elsewhere.

²⁵⁶ I recall no mention of specialized IA book trucks.

3.3 *Structural Residue*

By design, this dissertation has focused mainly on the early years of these large-scale information democratization efforts: what motivated them, how they defined their intended audiences and collections, and what they did and built in service of those motivations and definitions. This section will introduce a different perspective: the view from beyond the projects' end, looking both backward at the projects themselves and forward to their aftereffects. Of these four projects, only one – the Boston Public Library – remains extant and strong, with no signs of fading away. Two others – the Carnegie Library Program and the Open Content Alliance – no longer exist, and have not for some time. And the fourth, the Google Books Library Project, is now in its sunset years, as the company loses interest and scanning continues to slow. And yet, all of these projects have left – and will continue to leave – impressions on the structures through which we access information; durable traces of their existence that have meaning for our relationship with information resources. I refer to these traces as “structural residue:” a term which I take to refer mainly to tangible objects that outlive their social context of creation, and go on to have largely unintended consequences for subsequent social projects and arrangements. At the same time, however, I recognize that tangibility can be relative: one may not be able to physically touch a PDF or an eBook, but those format selections nonetheless structure information access in ways that can and do survive the contexts of in which they were chosen. A few examples from the cases examined here will help to further illuminate this concept.

As noted, the Boston Public Library is alive and well – and its central location is still half-housed in a building that was first opened in 1895 (a large addition, called “the Johnson Building” was appended to the original “McKim Building” in 1972 (Boston Public Library)). The original 1895 building remains beautiful and more or less functional more than a century later – and yet it has carried along with it across the years many pragmatic deficiencies from its original design. Indeed, a 1941 sourcebook on library architecture in America likened the McKim building to “an Italian palace...with a library fitted into it,” though quickly added that it was executed “with such loveliness of proportion and detail that its defects as a functional plan are overlooked” (Wheeler and Githens 1941).

Speaking as a sometime researcher in several parts of the older building, it is clear that in many areas, form preceded function in the design in ways that have been only partially ameliorated by later workarounds and renovations. For example, in the modern BPL, parts of many different vintages of catalogue coexist, from OPACs to card catalogs to microfiche, and the same is true of the types of stack access offered, from fully closed at the Delivery Desk, to partially closed in the Fine Arts and Government Documents departments, to fully open in the Bates Hall reading room and most of the Johnson Building. Indeed, the BPL is unlike any other library I have ever used in the prevalence and diversity of legacy systems still in active use – and certainly, this multiplicity of systems does not ease the task of the

information seeker, though for a researcher interested in the history of the library itself, it does provide a unique experiential window to several points in the library's past. And of course, all of these structures were at some point cutting-edge, the best of their historical moment (and some are even the best of the present moment); they were put in place for specific reasons, based on the information and options available at various points over the library's 120 years of history. But they have proven stubborn, persisting well beyond the point(s) when they seemed to represent the best option(s). The sentimental attachment to the McKim building itself is enormous, and for good reason: the structure is both beautiful and of deep historical importance, locally and beyond. And because of this, the building is not likely to go anywhere or undergo radical internal alterations any time soon. The persistence of other bits of structural residue – the catalogs, the stacks – can partially be chalked up to the structure of the building, whose architecture was intended to enable specific forms of access (e.g. closed stacks, pre-computer-era catalogs), though part of this persistence is undoubtedly more up to path dependence and a lack of sufficient funding to make complete updates. Taken as a whole, however, within the Boston Public Library's McKim building, a multitude of deeply political, personal, and historically situated design decisions, made mainly in the late 19th century, continue to retain structural sway over public information access in Boston to this day.

The Carnegie Corporation of New York ceased library donations nearly a full century ago, in 1919. Yet, before it ended, the program had funded the construction of 2509 libraries worldwide, which communities had to contend with in figuring out how to deal with later local information needs. And as Macleod notes, due to a combination of sentiment, inertia, and budgetary concerns, many of these buildings became anchors on progress, holding communities back from effectively serving the information needs of their local publics for decades beyond the end of the Carnegie donations:

Since the Carnegie structures were not about to crumble into dust, the temptation to remodel them was always severe, although their impressive facades hid numerous flaws. Small, awkward rooms made them prohibitively expensive to change. Floors were sometimes weak and the areas upstairs were too expensive to supervise. Carnegie's buildings had thus had an accidental effect opposite to what he intended: they choked off local initiative, since they were public monuments, too impressive to demolish. Not until the 1950's was there any meaningful movement to replace them, since towns with Carnegie buildings had not grown accustomed to arranging accommodations for their own libraries (Macleod 1968, 142).

This is a textbook example of a perverse consequence: Carnegie library buildings, donated out of a sincere desire to expand information access throughout the English speaking world, soon became impediments to the realization of that very goal. And this is an issue with which some recipient communities continue to contend: a survey made by George

Bobinski in 1992 found that of the 1681 buildings built with Carnegie money in the United States, 911 were still in use as public libraries at the time of the survey, with most of the others having been demolished or converted to other uses (Strum 1992).²⁵⁷ In the communities housing those 911 libraries (whose number is likely fewer now, two decades later), many difficult decisions have undoubtedly had to be made in order to preserve those buildings as functioning libraries – in particular, significant funds must have been dedicated to the preservation and repair of these buildings that could have gone to new construction, and many workarounds and legacy structures similar to those that persist in Boston have likely pervaded practices of information access within these buildings.

The buildings themselves, however, are not the only piece of structural residue that persisted long past the end of the Carnegie program itself: Bertram's six standardized library plans also had a significant afterlife. As van Slyck explains,

'Notes on the Erection of Library Buildings' was remarkably effective until World War II in promoting the open-plan library, with its centrally located charging desk and flanking reading rooms, as the ideal for small public libraries. In the prosperous 1920s, many communities were prompted to finance their own library buildings, often hiring architects whose plans revealed their previous experience designing Carnegie libraries. In the lean years of the 1930s, the Public Works Administration (PWA) filled the financial void left by the demise of the Carnegie library program, funding small, symmetrical, classically detailed libraries with open plans (Van Slyck 1995, 218).

The edicts laid out in Bertram's 'Notes' had achieved such pervasiveness within the Carnegie library program that they went on to influence library architecture for decades after that program had ended, further cementing the equation between what a typical *Carnegie* library *did* look like (at least in the later years of the program) and what a typical *public* library *should* look like: simple, classical, open-planned, and symmetrical, with every wall lined with freely accessible books.

The structural residue of the digitization projects, of course, has had less time to coalesce and accrete, as Google still continues to scan, and the Open Content Alliance – though a few years defunct as such – has largely dissolved into the Internet Archive's scanning initiatives, which are also ongoing. Nonetheless, it is perfectly possible to project at least a few elements of structure that will persist beyond the active lifespan of both projects. First, and most obviously, both of these projects have converted millions of books to digital form, and those files will live on, regardless of what happens to either Google or the Internet Archive. Why am I so sure? Because a central piece of both of these projects was a commitment to share those files with the libraries involved, and not only do libraries have

²⁵⁷ For example, the former Carnegie library near one of my old apartments, in Seattle's Ballard neighborhood, now houses a Pilates studio and an Australian-themed pub called the Kangaroo and Kiwi, among other tenants.

an excellent, multi-century track record of preserving cultural heritage materials, but the libraries involved with these projects have already begun to group together of their own accord to build independent, massive, and potentially sustainable collective repositories of their scanned files – in particular, HathiTrust and the Digital Public Library of America.

And yet, along with the (to my mind) undeniable good of ensuring the persistence of these files, several bits of less beneficial residue will persist along with them for a long time to come, in the form of imaging errors,²⁵⁸ idiosyncratic and unpredictable collection gaps introduced by scanner limitations and the condition of materials, and, of course, the epic and daunting number of problems with the metadata. All of these issues apply to both projects, though each one more to Google than to IA – and because Google has scanned at least four times as many books as IA, its problems will pervade and persist on a much broader scale. And when one considers that these scans are now being treated as a reliable or even apparently objective basis for the sorts of “big data” analyses that are now very much in vogue (e.g., Michel, et al. 2011), these structural issues have the potential to imperil not only the findability and readability of books for the average user, but the validity of pronouncements on social trends emerging from research in these corpora. This will be a major problem for scholarship going forward (although one wonders how many of those performing this type of research will recognize it as such), and has the potential to incur consequences equally as perverse as Carnegie libraries holding back information access: such computational linguistic analyses were, after all, one of the earliest hoped-for uses of the Google corpus, and yet the way in which that corpus was built and annotated may end up tainting or even invalidating the results of such research conducted within it.

3.4 *Physical vs. Intellectual Access: The Question of Guidance*

All of the three preceding themes have dealt with phenomena that were present in all or most of the cases examined here. The final theme, however, will explore an absence. That is, in all of these cases, the shape of the initiative has proceeded generally on a sort of *Field of Dreams* model, where “build it and they will come” becomes translated into “gather together great mountains of books, and the public will read (and improve themselves).” And yet, any librarian in the world would tell you that the provision of information resources is only the first step: users always, *always* need help. But still, in all four of these cases, the provision of such guidance has proven a much-delayed afterthought where it has not been ignored entirely.

In Boston, for example, the first iteration of the library opened to the public in 1854, and the first dedicated building in 1858; and yet, the first mention of a need for substantive patron

²⁵⁸ Google Books is notorious for the prevalence of scanner fingers and blurred pages in its imagery. As noted in Chapter 5, the company is working to algorithmically correct some of these errors, but one would expect that such efforts could extend over years, if not decades – assuming they don’t first lose interest and give up on trying.

assistance does not appear in the historical documentation of the library until the mid-1860s, when it was pointed out by Justin Winsor, acting as head of the library's Examining Committee ("Report of the Examining Committee" 1865). In the Carnegie case, a similar phenomenon occurs, though perhaps to a more radical extent: Carnegie donations, after all, only provided for physical library buildings: the provision of both service and the collections themselves were left up to the communities. And yet, the tax support required by the Carnegie program was patently insufficient to provide both collections and services in many of the smaller libraries donated – and those small libraries comprised the vast majority of Carnegie donations. In fact, as previously described, the CCNY ceased Carnegie's library donations in large part because of Johnson's finding that the lack of skilled guidance in many of these libraries was leading to their underutilization or even closure (Johnson 1916, Van Slyck 1995, 217, Jones 1997, 101).

And of course the digitization initiatives seem to be following this troubling precedent completely. Neither Google nor the Internet Archive provides more than superficial assistance to users. And although some of users' questions are getting filtered out to partner libraries whose names appear next to the books online (as I observed while working reference at the University of Michigan), many instances of user confusion are undoubtedly going unaddressed, imperiling the utility of these services for many of their public-educational aims. Going forward, as we digitize more, and establish new projects of this kind, this issue will only become more pervasive. Given the repeatedly proven fact that individuals will need assistance using huge swaths of informational materials such as those provided both by libraries and by digitization initiatives, whose responsibility is it to provide this guidance, if anyone, and how might that guidance best be provided? This will form a fruitful ground for future research, which I intend to pursue.

4. Limitations

It is important to recognize the limitations of this dissertation, particularly with regard to gaps in the data, the role of my own subjectivity in interpretation, and the incomplete comparability of the cases chosen.

4.1 *Data Lacunae*

The largest limitation on the reliability of this analysis, across all four cases, emanates from gaps in the available data. For the historical cases, these gaps occurred for two basic reasons, one entirely out of my control, and one not. First, many potentially useful documents simply did not exist – or did not exist anymore. This, of course, is a generalized concern in all historical research: not all information useful to the historian is documented, not all documented information is kept, and not all kept documentation is retained in such a way as to allow later access by researchers. For example, the Carnegian focus on efficiency led to the discarding or returning-to-sender of many documents valuable to the history of

Carnegie's library program – and those that were centrally retained in the short run were also eventually microfilmed, and the paper versions then discarded. As such, documents that are illegible in the microfilm (of which there are hundreds, if not thousands) are essentially lost to history, along with whatever unknown and unknowable documents had been thrown away without ever having been filmed. And in the Boston case, I had initially identified several individuals whose correspondence I wanted to find – and especially their correspondence among one another: for example, Ticknor's correspondence with Everett. But despite poring over finding aids at the Boston Public Library and at Harvard (where both men also worked, and where Everett's papers are deposited), I was unable to locate a trace of this correspondence for many of the individuals I had pre-identified. It is possible, of course, that no such correspondence ever existed – though I think it more likely that it has either been discarded or is located in a place where I did not think to look.

And in fact, that leads directly into the historical data limitation that was partially under my control: the selection of archives to scour. The boundaries on that selection had to do mainly with costs. It is quite expensive to take cross-country trips to visit archives, and so only a select few could be targeted: the Boston Public Library's internal archives, the CCNY Archives at Columbia University, and the Andrew Carnegie Collection at the Library of Congress. Indeed, the high cost of such excursions is one of the reasons that digitization is so valuable and transformative, especially for historical research like this – and because of prior digitization, and especially that which had already been executed by Google and the Internet Archive, I was also able to supplement the documentation located on my archival visits with materials scanned from other libraries and archives worldwide. Despite these efforts, however, there are surely many materials that I have missed. For example, although I had initially hoped to explore the institutional archives of one or more Carnegie-donated libraries, and although such documents would likely have enriched the analysis of that case, I never had the opportunity to make such a visit. As with the discarded or nonexistent documentation discussed above, it is difficult to know what might have fallen into the gaps left between collections – though one suspects the perspective of the Carnegie recipient communities would be one victim of such a fall.

Within the LSDI cases, gaps in the data have partially similar and partially different origins. Similar to the historical cases, not all documentation of potential use to this dissertation was kept – by Google, by the Internet Archive, or by any library partner. Many interviewees, when asked if there was any supplementary documentation they might provide, indicated that most of it was in email, and that email had since been deleted.²⁵⁹ Beyond this, however, for those involved with the GBLP, non-disclosure agreements also intervened. The

²⁵⁹ As an aside, this type of practice gives me night-terrors for the historians of the future. Where will those historians find the treasure troves of correspondence like those we can examine from the nineteenth century, to illuminate what's happening now? Largely, they won't. And that is an inestimable loss.

Googlers did not feel they could provide any documentation beyond what was already publicly available on the Google website, and most of the librarians did not either. All were covered by non-disclosure agreements, and yet, especially among the librarians, there seemed to be both a lack of clarity and a distinct nervousness about what was actually covered by those agreements – with the few librarians who did provide internal documents being the rare exceptions to that rule.

For the OCA, although no non-disclosure agreements existed, there seems to have been less documentation to (non-)disclose. Where Google's project seems to have been heavily internally documented (at least, after a point (P12)) and also clearly defined by many contracts and agreements, the OCA was much looser, and very little seems to have been written down, whether for internal or external use.

Finally, the data for both of the LSDI cases are limited by the set of interviewees I was able to gather – and the set that I was not. For most institutions involved, I was only able to procure one interview, both because of time limitations and because not all of those contacted responded, or were willing to be interviewed. Thus, although I made my best effort to interview the individuals indicated as especially influential by others involved, I was not able to do so for many individuals so identified. For example, although I was able to talk with the central figure in starting the OCA, Brewster Kahle, I was not able to talk with the central figure in starting the Google Books Library Project, Larry Page. That specific example, however, illustrates a further difficulty in gaining participation from the type of individuals in the population relevant to this study: that is, they are generally “elites” – library directors or associate directors, leaders or managers within technology companies, and others with multitudinous responsibilities and not much incentive to or necessary interest in participating in some unknown-to-them doctoral student's dissertation research. Indeed, the fact that I was able to procure the 18 interviews that I did obtain has a great deal to do with my personal history and the network it allowed me to build; I doubt I could have successfully solicited the participation of nearly as many of those individuals had I not already crossed paths with many of them through my work on PR for the Google Books Library Partnership at the University of Michigan. However, the personal and situated nature of that network is itself a potential limitation – it is very possible that I had more success recruiting individuals with particular views on both projects because I had participated in the UM side of things. Still, I did my best to ensure that a range of perspectives was included; although additional interviews likely would have deepened the insights provided – or potentially shifted them somewhat – my own available resources and the relative “recruitability” of potential participants ended up limiting the pool.

4.2 Subjectivity

The issue of personal involvement noted above also relates to the second limitation of this research. That is, as a piece of qualitative work with a single interviewer and, for the most part, a single coder of the data, a great deal of this analysis emanates from my brain, with all its internal biases and preconceptions. I have sought throughout to let the data speak through me, rather than impressing my own preconceived notions upon it, but I admit that especially for the digitization cases, that has sometimes been difficult, and I may not always have been completely successful in taking a balanced view. For all of these cases, I admit that I am inclined to try to see the positive: I truly believe, personally, that more information access of this kind is good for the world, and I am thus predisposed to respect and admire initiatives that promote such access, especially in such radically ambitious ways. This has more than likely blunted my criticism of each case at various points, for better or worse. And particularly with regard to Google Books, my previous involvement, especially as a person tasked with arguing in favor of the project, has undoubtedly influenced my interpretation of the data surrounding that case. Still, I would note that while so much that I have read about that project has erred toward extremes – Google Books as the devil or Google Books as the savior of all mankind – I believe my analysis has avoided either one, recognizing both the project’s significant merits and its serious flaws (as indeed I have striven to do for all the cases presented here).

Additionally, subjectivity and bias are also relevant on one other level: that is, in the statements of the interviewees. Memories are not perfect, and humans have a near-infinite capacity for rationalization of actions after the fact; as such, the interview accounts are undoubtedly colored to some extent by the time that has passed since the events described actually occurred. For example, early positive experiences with the IA or with Google by a partner librarian could have been forgotten or diminished in memory by later negative ones, or vice-versa. Or, for motivations, the passage of time could lead to the papering over of the actual driving forces behind the projects with after-the-fact rationalizations and/or stories that had by then become canonical, like that of the idea for Google Books preceding the idea for Google itself. Throughout the analysis, triangulation with other interviewee perspectives as well as secondary sources has been employed to ameliorate these biases. Nonetheless, it is likely that at least some such bias, specific to the individuals who participated, is still reflected here.

4.3 Comparability

Finally, as in any comparative case study analysis, these cases are not 100% parallel. In fact, in this research, the percentage would have to be placed far lower, especially for the comparability of the nineteenth century cases to the twenty-first century ones: the historical context is simply too significant to be ignored. For all of these cases, the social context – which includes the historical context – is crucially important to the way in which the project

was defined and developed. And yet, that macrosocial context has not been an explicit focus of this analysis, except to the extent that it has been reflected in the actions or attitudes of specific individuals who had an influence on each project. This limitation, of course, is largely a matter of project scope. This project was already very large, and adding an examination of macrosocial trends surrounding the historical development of each case would have made it many times larger. Still, such an examination could be a very interesting direction for future research.

5. Looking Forward: The Future of the [Universal] Library

The dream of the universal library is a compelling one, and one that I believe is worth continuing to strive toward, even though there is a vanishingly slim chance we will ever achieve it. Universality inevitably slips beyond our grasp, for we are limited. We cannot imagine all the uses to which information might be put, or all the users who might have an interest in what we've brought together. And any attempt to assemble the universal collection – whether of books or of any other sort of media – inevitably devolves rapidly into a problem of philosophy: if we seek all the books, how are we defining “book”? And what would it mean to find “all” of them? And even if we can find answers to both of these questions, how would it ever be pragmatically possible to locate “all” the “books” that have ever been published in the history of the world? It seems clear to me that the answer is that we cannot. And with regard to serving “everyone,” there is no such thing as a universally accessible technological system. Any system designed by humans will reflect some of the attitudes and goals of the humans doing the design, and those social elements will end up making the system more accessible to some, and less accessible to others – though not necessarily in ways intended by the designers (often decidedly not).

Still, the projects we embark upon in order to strive toward these utopian visions of universality can and do have amazing positive benefits. There is little question, for example, that all four of the cases examined here have put information resources into the hands of individuals who would not otherwise have had access to them. And while there are certainly those in this world who believe that more information in the hands of the public (or at least, all segments of the public) is not necessarily a good thing – witness the so-called “Great Firewall of China,” or the ongoing violent clashes over educating women in various parts of the world – these opposing forces only provide further substantiation for the argument that access to information is empowering: after all, why else would totalitarian regimes seek to restrict it? And if information is empowering, and these types of initiative provide mass quantities of information to large numbers of people, then it would follow that these types of project are – or have the potential to be – significant forces for empowerment through the equalization of educational opportunity within their respective spheres of influence – just as they were always intended to be.

Moreover, the digitization cases have arguably had some other public benefits, beyond the expansion in access to information that they have facilitated. In particular, where before Google Books, libraries questioned whether it would even be possible or useful to scan entire collections, now that conversation has entirely changed: it is no longer a question of “whether,” but “how,” “when,” and “with whom.” Additionally, the lawsuit against Google (along with the successor case against HathiTrust) has also had some pro-social outcomes, in the sense that two courts have now judged wholesale scanning to be a transformative fair use – and as I have argued elsewhere (Jones 2013c), whatever one thinks of Google or its book-scanning, the precedent set by these fair use rulings will have major benefits for libraries and others hoping to embark on digitization, especially as copyright terms grow longer and longer even as twentieth-century media continues to decay.

At the same time, however, it is worth pondering the top-down nature of these projects for a moment, and what it might mean for their potential to empower the public. In none of these cases was the public actively and routinely consulted in the process of shaping the information access initiative produced; rather, to the extent that public voices were present in design at all, they were highly mediated, through several levels of other stakeholder groups. And in fact, in terms of vision, a huge portion of each of these projects depended on the personal predilections of a single wealthy individual – hardly a democratic process for producing a theoretically democratizing product. Is this kind of origin story necessary for initiating such grandiose initiatives?²⁶⁰ Or is this commonality among the cases merely an accident of history, or perhaps case selection? The fact that most of these projects (certainly the last three) were initiated during periods of enormous wealth disparity seems relevant also: at what point do we begin to rely more heavily on the personal largesse of the wealthy than on the collective action of public bodies?

Beyond the point of initiation, moreover, there is a significant divergence between the funding strategies of the public libraries and those of the LSDIs: namely, the founders of public libraries – Boston Public, all of the Carnegie libraries, and beyond – always intended for their ongoing funding to be provided by local taxation. And on the whole, this has proven quite a sustainable model. While current economic conditions and shifting local priorities may now be imperiling this model in many regions, it functioned reasonably well for decades before that. The LSDIs, on the other hand, had no planned source for ongoing funding at the point of their initiation, beyond, perhaps, hope. Both the GBLP and OCA seem to have been founded with the sense that new pots of money would simply continue

²⁶⁰ It is worth noting that at least one digitization initiative – Project Gutenberg – has taken the inverse approach, starting at the grass roots and building from the bottom up. However, it is equally worth noting that this approach has never been able to achieve anywhere near the scale of the projects described here: in almost 40 years, Project Gutenberg has only made about 100,000 books digitally available. It is miniscule in comparison to the two LSDIs examined here.

to appear, almost magically, like coins in a Super Mario game. And yet, hope is not a strategy, and neither is magic. When Microsoft pulled its funding from the OCA, it kneecapped the initiative, and the OCA didn't survive. And now that Google is dialing down its involvement (and investment) in scanning, no other major players appear to be stepping up to carry its financial torch either. On balance, it seems that the future funding of both these book scanning efforts and the digital outputs they have already produced will rest with libraries – and specifically, with major research libraries, as those are the ones who have participated in these projects. The Internet Archive's ongoing model for scanning provides one glimpse of how this might work: libraries pay IA to scan their materials at a lower rate and higher pace than the library would be able to achieve using its internal resources, and those materials generally end up deposited with the IA as well as the library for ongoing custodianship. And in terms of storage and ongoing access, HathiTrust's model of soliciting member libraries beyond those from whom the bulk of materials have been scanned, and sustaining itself on those libraries' membership fees, seems plausibly sustainable as well. And yet, this is quite a major shift in role for these academic libraries. The average college or university library, after all, has the central mission of serving the college or university of which it is a part. If the public is even allowed in the door, there is often some sense that the university is doing them a favor – serving the public goes beyond these libraries' central mission. Yet, it is these same academic libraries that are being tasked with funding the continuation of scanning and hosting digital books *for public consumption*. And as more and more such libraries take on these public-facing responsibilities, it may behoove them to reexamine their underlying missions, to see where these novel initiatives might fit in with and potentially complement existing institutional aims.

The truly universal library may remain consigned to the realm of fiction, but Raber's "library faith" persists nonetheless. The belief that providing more information to more people will lead to a better world – one that is more egalitarian, more democratic, and more enlightened – still holds great sway. It continues to propel individuals to create fantastical, often initially impossible-seeming initiatives, in hopes of bringing that better world into being. And that, I think, is a marvelous historical trend, and I hope it continues. At the same time, however, it is important that this belief does not turn into *blind* faith; that we remain conscious of what exactly we are trying to do when building these utopian information initiatives, and for whom we are doing it, and why. Especially as information access systems gain ever-greater scale and ever-more-global spread, it bears keeping in mind that technological systems are not objective or neutral, but rather carry with them embedded traces of the social contexts of their creation – including the conscious or unconscious attitudes and biases of their creators. This dissertation has sought to expose the social embeddedness of four massive and influential information democratization efforts, and to reveal some of the ways in which a balanced understanding of the foci and blind spots of

past information access initiatives might be able to help the designers of present and future information access initiatives make better decisions: ones that break down more barriers among users, create more welcoming spaces for discovery and innovative thinking, and enable more effective access for all. Universality may indeed prove elusive, but we continue to try nonetheless. It is my hope that this dissertation will help guide the way toward doing so in a more mindful way.

Appendix A: Coding Framework

Research Questions:

- **RQ1: Motivations (WHY):** Statements that indicate motivations for pursuing an initiative of the kind described in RQ2
- **RQ2: Defining the Initiative (WHAT):** Statements that indicate how the initiative (or its leadership) defines what it is doing, in terms of use and collections.
 - **Users:** Statements that describe the intended user base of the initiative
 - **Uses:** Statements that describe how leadership thought people would use the thing they were creating
 - **Users: Composition:** Statements that describe the intended user base of the initiative in terms of composition - what kinds of people did they think/hope/believe would use it?
 - **Users: Size:** Statements that describe the intended user base of the initiative in terms of size - how many people did they think/hope/believe would use it?
 - **Collections:** Statements that describe the shape of the initiative's collection
 - **Collection: Composition:** Statements that describe the shape of the initiative's collection in terms of what kinds of things it would hope/strive to include - or exclude
 - **Collection: Size:** Statements that describe the shape of the initiative's collection in terms of how large it would be - comprehensiveness a goal?
- **RQ3: Achieving the definitions in RQ2 (HOW):** Statements that indicate how the initiative went/is going about achieving the definitions described in RQ1.
 - **Structures:** Structures put in place in order to help achieve the definitions described in RQ1 (design decisions)
 - **Interface:** Interface elements (pieces of the design directly touched by users)
 - **Backend:** Backend features (pieces of design hidden from users)
 - **Processes:** Processes or procedures put in place in order to help achieve the definitions described in RQ1 (procedural or regulative decisions or moves)
 - **Access:** Processes/procedures related to the provision of access to the books/materials.
 - **Administration:** Processes/procedures put in place regarding administration of the institution/collaboration.
 - **Costs:** Statements that indicate how the initiatives went about covering the cost of what they wanted to do
 - **Upfront:** Statements about upfront costs; initial investments
 - **Ongoing:** Statements about ongoing costs; sustainable (or at least forward-going) funding models

Theory:

- **Sociotechnical Cycle:** Statements about the mutual influence between social and technical systems
 - **Social influencing technical:** Statements about social concerns influencing the design of a technical structure or system (incl. architecture)
 - **Technical influencing social:** Statements about technical capabilities influencing what is or ought to be done in social settings

- **Power:** Statements illustrating a power dynamic or inequality at play
 - **Inequality of Influence:** Passage that indicates that some individual or group has more or less influence than another individual or group
 - **Mediation:** Passage that indicates that the influence of some individual or group has been mediated through another individual or group
 - **Transparency:** Passage that indicates that transparency or lack of transparency served to create power differentials
- **Information:** Passage that indicates that the informational content of these structures gives them additional social valence for power structures
- **Scale:** Passages indicating a direct influence of the scale of the project. That is, ways in which things had to happen differently because of the large scale.

Appendix B: Primary Source Documents

Boston Public Library

Annual Reports:

- "First Annual Report of the Trustees of the City Library." *City Document No. 73*. City of Boston. November 14, 1853.
- "Second Annual Report of the Trustees of the City Library." *City Document No. 74*. City of Boston. November 2, 1854.
- "Third Annual Report of the Trustees of the City Library." *City Document No. 57*. City of Boston. November 15, 1855.
- "Fourth Annual Report of the Trustees of the City Library." *City Document No. 61*. City of Boston. October 31, 1856.
- "Fifth Annual Report of the Trustees of the City Library." *City Document No. 71*. City of Boston. November 10, 1857.
- "Sixth Annual Report of the Trustees of the City Library." *City Document No. 46*. City of Boston. November 15, 1858.
- "Seventh Annual Report of the Trustees of the City Library." *City Document No. 66*. City of Boston. November 10, 1859.
- "Eighth Annual Report of the Trustees of the City Library." *City Document No. 90*. City of Boston. November 19, 1860.

*Bates, Joshua. Letter to Thomas W. Ward, October 1, 1852.

— — —. Letters to Samuel G. Ward, 1853-64.

Board of Aldermen. "An Ordinance in Relation to the Public Library Commissioners." *City Document No. 19*. March 5, 1855.

— — —. "Communication from the Mayor, Transmitting a Memorial from the Trustees of the Public Library, Suggesting Certain Changes in the Library Ordinance." *City Document No. 70*. November 5, 1857.

Board of Trustees. "Rules and Regulations of the Public Library of the City of Boston [Annotated by G. Ticknor]." August 17, 1858.

Boston Common Council. "Report of Committee on Public Library." *City Document No. 79*, 1852.

— — —. "An Ordinance in Relation to the Public Library." *City Document No. 57*. October 14, 1852.

— — —. "Report of Committee on Public Library." *City Document No. 21*. February 23, 1854.

— — —. "An Ordinance in Relation to the Public Library Commissioners." *City Document No. 120*. December 21, 1854.

— — —. "Public Library 1855." *City Document No. 46*. June 6, 1855.

*Capen, Edward. "Reports of the Librarian." 1853-68.

*— — —. "Sketch of the Early History of the Boston Public Library." 1900?

*"Documents Relating to a Meeting of Citizens of Boston Held May 5, 1841, to Consider Mr. A. Vattemare's Plan of a Public Literary and Scientific Institution [in Boston] and System of International Literary Exchange." Boston, 1841.

*Greenough, William W. "Annals of the Public Library of the City of Boston from the Year 1848 to Its Removal to the New Edifice in Copley Square." Boston, MA: Boston Public

Library: Rare Books & Manuscripts, 1891?

"Message of the Mayor on the Subject of a Public Library." *City Document No. 10*, 1852.

**Proceedings of the Trustees*. Vol. 1. Boston: Public Library of the City of Boston, 1852-1858.

**Proceedings of the Trustees*. Vol. 2. Boston: Public Library of the City of Boston, 1859-1863.

Proposed Rules and Regulations of the Public Library of the City of Boston. Boston: J.H. Eastburn, City Printer, 1853.

**Quincy, Josiah Phillips. *Appeal in Behalf of the Boston Athenaeum*. Boston: John Wilson & Son, 1853.

"Report of the Committee on the Library, in Relation to the Donations Received from the City of Paris, with a Catalogue of the Reciprocal Gifts Exchanged between the Two Cities, with the Names of the Donors; Together with the Proceedings of the City Government Upon the Subject of International Exchanges." *City Document No. 46*. September, 1849.

"Report of the Trustees of the Public Library of the City of Boston." *City Document No. 37*. Boston City Council. July 6, 1852.

Reports of the Examining Committees, 1853-68. Boston Public Library, 1853-68.

Specifications for Shelving the Public Library Building. Boston: George C. Rand and Avery, 1857.

State of Massachusetts. "An Act to Authorize the City of Boston to Establish a Public Library." 1848. Retrieved February 29, 2012 from:
<http://archives.lib.state.ma.us/handle/2452/94428>.

State of Massachusetts. "An Act to Authorize the City of Boston, to Found and Maintain a Public Library." 1853. Retrieved February 29, 2012 from:
<http://archives.lib.state.ma.us/handle/2452/95985>.

Ticknor, George. *Union of the Boston Atheneum and the Public Library*. Boston: Dutton and Wentworth, 1853.

**Trustees of the Public Library. *Memorial of Joshua Bates*. Boston, 1865.

Whitney, James Lyman. *Reminiscences of an Old Librarian*. Boston, 1909. Retrieved June 10, 2011 from The Internet Archive.

Winthrop, Robert Charles. *Proceedings on the Occasion of Laying the Corner Stone of the Public Library (Address by R.C. Winthrop)*. Boston, 1855.

Winthrop, Robert C., Samuel G. Ward, George Ticknor, Nathaniel B. Shurtleff, George Odiorne, and George W. Warren. [Advertisement]. *The Boston Daily Atlas*, February 17, 1855, [1].

Carnegie Libraries

Annual Reports:

Allegheny, PA:

"First Annual Report of the Carnegie Free Library, Allegheny, Pa, for the Year Ending 28th of February, 1891." Derrick Publishing Company, Printers and Bookbinders. February 28, 1891. Retrieved March 22, 2012 from HathiTrust.

Stevenson, William M. "Second Annual Report of the Carnegie Free Library, Allegheny, Pa, for the Year Ending February 29th, 1892." Press of John Ogden. February 29, 1892. Retrieved March 22, 2012 from HathiTrust.

"Third Annual Report of the Librarian of Carnegie Free Library, Allegheny, Pa, for the Year Ending 28th February, 1893." February 28, 1893. Retrieved March 22, 2012 from

- HathiTrust.
- Atlanta, GA:
- "First Annual Report of the Board of Trustees of the Carnegie Library of Atlanta." Pease Printing Co., 1900. Retrieved March 22, 2012 from Google Books.
 - "Second Annual Report of the Board of Trustees of the Carnegie Library of Atlanta." 1901. Retrieved March 22, 2012 from Google Books.
 - "Third Annual Report, Carnegie Library of Atlanta." Mutual Publishing Co., 1902. Retrieved March 22, 2012 from Google Books.
- Braddock, PA:
- "Twentieth Annual Report of Carnegie Free Library and Sixteenth Annual Report of Carnegie Club, Braddock, Pa, for the Year 1908, Together with Statistics from the Founding of the Library." 1908. Retrieved March 22, 2012 from HathiTrust.
- Pittsburgh, PA:
- Carnegie Library of Pittsburgh. *First Annual Report to the Board of Trustees of the Carnegie Library of Pittsburgh*: Carnegie Library of Pittsburgh, 1897:
<http://books.google.com/books?id=MQ8bAAAAMAAJ>.
 - — —. *Second Annual Report to the Board of Trustees of the Carnegie Library of Pittsburgh*: Carnegie Library of Pittsburgh, 1898:
<http://books.google.com/books?id=MQ8bAAAAMAAJ>.
 - — —. *Third Annual Report to the Board of Trustees of the Carnegie Library of Pittsburgh*: Carnegie Library of Pittsburgh, 1899:
<http://books.google.com/books?id=MQ8bAAAAMAAJ>.
- Bertram, James. "Notes on the Erection of Library Buildings." 1915. Appendix in *Free to All: Carnegie Libraries & American Culture, 1890-1920*, by Abigail Van Slyck. Chicago: University of Chicago Press, 1995.
- Carnegie, Andrew. "Atlanta." [Speech], 1909. Andrew Carnegie Papers, Box 254. Courtesy Library of Congress Manuscript Division.
- — —. *Autobiography of Andrew Carnegie*. London: Constable & Co. Limited, 1920. Retrieved March 19, 2012 from Project Gutenberg.
 - — —. "The Best Fields for Philanthropy." *The North American Review* 149, no. 397 (1889): 682-98. Retrieved July 8, 2009 from JSTOR.
 - — —. "Brooklyn." [Speech], 1909. Andrew Carnegie Papers, Box 254. Courtesy Library of Congress Manuscript Division.
 - — —. [Dedication of Jedburgh Library Building]. October, 1894. Andrew Carnegie Papers, Box 250. Courtesy Library of Congress Manuscript Division.
 - — —. [Founder's Day Address, Pittsburgh], 1898. Andrew Carnegie Papers, Box 251. Courtesy Library of Congress Manuscript Division.
 - — —. "Homestead Address." 1898. Andrew Carnegie Papers, Box 251. Courtesy Library of Congress Manuscript Division.
 - — —. "Presentation of the Carnegie Library to the People of Pittsburgh, with a Description of the Dedicatory Exercises." November 5, 1895. Retrieved August 14, 2012 from:
<http://www.carnegielibrary.org/research/pittsburgh/Carnegie/mrac1.html>.
 - — —. "Presentation of the Freedom of the Burgh of Inverness to Andrew Carnegie, Esq." September 12, 1890. Andrew Carnegie Papers, Box 250. Courtesy Library of Congress

- Manuscript Division.
- — —. "Royal Burgh of Wick, Caithness, Scotland. Presentation of the Freedom of the Burgh to Mr. Andrew Carnegie." August 4, 1890. Andrew Carnegie Papers, Box 250. Courtesy Library of Congress Manuscript Division.
 - — —. [Speech delivered in Montgomery, Alabama], 1909. Andrew Carnegie Papers, Box 254. Courtesy Library of Congress Manuscript Division.
 - — —. "Wealth." *The North American Review* 148, no. 391 (1889): 653-64. Retrieved July 8, 2009 from JSTOR.
 - — —. "Address by Mr. Carnegie." In *Souvenir: Laying the Foundation Stone of the Peterhead Free Library and Museum*. August 8, 1891. Andrew Carnegie Papers, Box 250. Courtesy Library of Congress Manuscript Division.
- Carnegie Public Library, Ayr, and Andrew Carnegie. "Address by Andrew Carnegie, Esq., on the Occasion of Laying the Memorial Stone." October 5, 1892. Andrew Carnegie Papers, Box 250. Courtesy Library of Congress Manuscript Division.
- *Correspondence, 1890-1916. Andrew Carnegie Collection. [Search Term: "Carnegie Library"; Manual filtering to correspondence related to the library program]. Retrieved October 19, 2011 from: <http://diva.library.cmu.edu/webapp/carnegie/index.jsp>. Courtesy Carnegie Mellon University Libraries, Digital Collections.
- "Dedication of Washington (D.C.) Carnegie Library Building." *Library Journal* 28, no. 1 (January 1903): 18: http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&res_dat=xri:c19index-us&rft_dat=xri:c19index:PCI:7614-1903-028-01-000015.
- **Free Public Library: Johnstown, Pennsylvania." In *Carnegie Corporation of New York Library Correspondence (Microfilm)*: Columbia University Rare Books and Manuscripts Library, 1890.
- **Free Public Library: Pittsburgh, Pennsylvania." In *Carnegie Corporation of New York Library Correspondence (Microfilm)*: Columbia University Rare Books and Manuscripts Library, 1890.
- **Free Public Library: Fairfield, Iowa." In *Carnegie Corporation of New York Library Correspondence (Microfilm)*: Columbia University Rare Books and Manuscripts Library, 1892.
- **Free Public Library: Braddock, Pennsylvania." In *Carnegie Corporation of New York Library Correspondence (Microfilm)*: Columbia University Rare Books and Manuscripts Library, 1895.
- **Free Public Library: Atlanta, Georgia." In *Carnegie Corporation of New York Library Correspondence (Microfilm)*: Columbia University Rare Books and Manuscripts Library, 1898.
- **Free Public Library: Carnegie, Pennsylvania." In *Carnegie Corporation of New York Library Correspondence (Microfilm)*: Columbia University Rare Books and Manuscripts Library, 1898.
- **Free Public Library: Pittsburg, Texas." In *Carnegie Corporation of New York Library Correspondence (Microfilm)*: Columbia University Rare Books and Manuscripts Library, 1898.
- Hendrick, Burton. Interview with James Bertram, for Andrew Carnegie's Biography. September 24-28, 1927. Andrew Carnegie Papers, Box 257. Courtesy Library of Congress Manuscript Division.
- "Mr. Andrew Carnegie in Aberdeen. Opening of the Public Library. Presentation of the

Freedom of the City." In *Souvenir: Laying the Foundation Stone of the Peterhead Free Library and Museum*. July 6, 1892. Andrew Carnegie Papers, Box 250. Courtesy Library of Congress Manuscript Division.

"Mr. Andrew Carnegie on Public Libraries: Opening of Auchterarder Institute." *The Library* s1-8(1) (1896): 501-04. Retrieved August 31, 2012 from Oxford Journals Online: <http://library.oxfordjournals.org/>.

Google Books

Interviews: P2, P3, P4, P8, P9, P10, P12, P13, P16, P17, & P18

Armstrong, Kim. "One Million Books Scanned and Returned to CIC University Libraries." *Inside Google Books* (blog). February 4, 2011. Accessed February 4, 2011. <http://booksearch.blogspot.com/2011/02/one-million-books-scanned-and-returned.html>.

Barnett, Doug, and Travis Willmann. "The University of Texas Libraries Partner with Google to Digitize Books." January 19, 2007. Retrieved January 17, 2012 from: <https://www.lib.utexas.edu/about/news/google/>.

Biblioteca Complutense. "Búsqueda De Libros Electrónicos UCM-Google." <http://www.ucm.es/BUCM/atencion/25403.php>. Accessed April 27, 2011.

— — —. "Proyecto De Digitalización Biblioteca Complutense-Google." <http://www.ucm.es/BUCM/biblioteca/11979.php>. Accessed April 27, 2011.

Bibliothèque Cantonale et Universitaire – Lausanne. "La Bcu Et Google Livres ". <http://www.unil.ch/bcu/page74263.html>. Accessed January 17, 2012.

Carr, Reg. "From Gutenberg to Google: The Case of the Bodleian Library, Oxford." 2005. <http://www.bodley.ox.ac.uk/librarian/rpc/gg2gg/gg2gg.htm>. Accessed April 27, 2011.

— — —. "The Oxford-Google Mass-Digitisation Project: How, Why and What?". Educause Webcast. June 15, 2005: <http://www.bodley.ox.ac.uk/librarian/rpc/educause/educauseGoogle.htm>.

— — —. "The Oxford-Google Mass-Digitisation Programme." *CNI Spring 2005 Task Force Meeting*, 2005. <http://www.bodley.ox.ac.uk/librarian/rpc/CNIGoogle/CNIGoogle.htm>. Accessed.

Cliatt, Cass. "Library Joins Google Project to Make Books Available Online." *News at Princeton*, Princeton University, 2007. <http://www.princeton.edu/main/news/archive/S16/84/71S02/index.xml>. Accessed April 27, 2011.

Coleman, Mary Sue. "Google, the Khmer Rouge, and the Public Good." Address to the Professional/Scholarly Publishing Division of the Association of American Publishers. February 6, 2006. Retrieved April 12, 2006 from: <http://www.umich.edu/pres/speech/archive/060206google.php>.

Columbia University Libraries. "Columbia University Libraries Becomes Newest Partner in Google Book Search Library Project." *Library News*, 2007. http://library.columbia.edu/news/libraries/2007/20071213_google.print.html. Accessed April 27, 2011.

— — —. "Frequently Asked Questions About the Columbia University - Google Book Search Library Partnership." *Library News*, 2007.

- http://library.columbia.edu/news/libraries/2007/20071213_googlefaq.print.html. Accessed.
- Colvin, Jennifer. "UC Libraries Partner with Google to Digitize Books." California Digital Library, 2006. <http://www.universityofcalifornia.edu/news/2006/aug09.html>. Accessed April 27, 2011.
- Committee on Institutional Cooperation. "Google Book Search Project - Introduction." *CIC - Google Book Search Project*. <http://www.cic.net/Home/Projects/Library/BookSearch/Introduction.aspx>. Accessed April 27, 2011.
- — —. "CIC-Google FAQ." *CIC - Google Book Search Project*, 2008. <http://www.cic.net/Home/Projects/Library/BookSearch/FAQ.aspx>. Accessed April 27, 2011.
- — —. "CIC-Google Government Documents Project." *CIC - Google Book Search Project*, 2010. <http://www.cic.net/Home/Projects/Library/BookSearch/GovDocs.aspx>. Accessed April 27, 2011.
- Committee on Institutional Cooperation, and Google, Inc. "Cooperative Agreement." *CIC - Google Book Search Project*. <http://www.cic.net/Home/Projects/Library/BookSearch/CIC-Google.aspx>. Accessed April 27, 2011.
- Cornell University Library. "FAQ: Cornell University Library-Google Library Partnership." *Library Communications*. <http://www.library.cornell.edu/communications/Google/faq.html>. Accessed April 27, 2011.
- Cunningham, Kelly Eileen. "Major Universities See Promise in Google Book Search Settlement." October 28, 2008. Retrieved January 20, 2012 from: <http://ns.umich.edu/new/releases/6807>.
- Ghent University Library. "Collections 1797; Ghent University 1817." May 23, 2007: <http://lib.ugent.be/files/en/pdf/20070523-persmap-en.pdf>.
- — —. "Google and Ghent University Library to Make Hundreds of Thousands of Dutch and French Books Available Online." 2007. <http://lib.ugent.be/files/en/pdf/20070523-persnota-en.pdf>. Accessed April 27, 2011.
- — —. "Ghent - Google - Treasures." August 28, 2009. Retrieved April 27, 2011 from: <http://lib.ugent.be/files/en/pdf/20090828-persmap-en.pdf>.
- Giersberg, Dagmar. "The Bavarian State Library's Books Are Being Digitalised by Google." *Goethe-Institut*, 2007. <http://www.goethe.de/wis/bib/dib/en2332286.htm>. Accessed April 27, 2011.
- Google, Inc. "Google Checks out Library Books." 2004. http://www.google.com/press/pressrel/print_library.html. Accessed June 8, 2008.
- — —. "Google Books Library Project - an Enhanced Card Catalog of the World's Books." 2011. <http://books.google.com/googlebooks/library.html>. Accessed March 7, 2011.
- — —. "Google Books Terms of Service." 2011. <http://books.google.com/intl/en/googlebooks/tos.html>. Accessed May 1, 2012.
- — —. "What You'll See When You Search on Google Books." 2011. <http://books.google.com/intl/en/googlebooks/screenshots.html>. Accessed November 11, 2011.

- Harvard University Library. "Harvard's Digitization Project with Google: Harvard Libraries and Google Announce Pilot Digitization Project with Potential Benefits to Scholars Worldwide." *Harvard University Library News Archive*, 2004. http://hul.harvard.edu/news/2004_1214_news.html. Accessed April 27, 2011.
- — —. "Online Now – Google Book Search for Harvard." *Harvard University Library News Archive*, 2008. http://hul.harvard.edu/news/2008_0307.html. Accessed April 27, 2011.
- Keio University. "Keio University to Partner with Google, Inc. For Digitalization and Release of Its Library Collection to the World for 'Formation of Knowledge of the Digital Era'." 2007. http://www.keio.ac.jp/english/press_release/070706e.pdf. Accessed April 27, 2011.
- Keller, Michael. "Stanford Initial Statement on Decision on Amended Settlement Agreement of the Google Library Project Suit." *SU LAIR*, Stanford University, 2011. <https://lib.stanford.edu/google-books/stanford-initial-statement-decision-amended-settlement-agreement-google-library-project>. Accessed April 27, 2011.
- Marsh, Ellen. "Cornell University Library Becomes Newest Partner in Google Book Search Library Project." *Library Communications*, 2007. <http://www.library.cornell.edu/communications/Google/>. Accessed April 27, 2011.
- "More About Google Books." *SU LAIR*, Stanford University, 2011. <http://lib.stanford.edu/google-books/more-about-google-books>. Accessed April 27, 2011.
- New York Public Library. "Google Books at NYPL." *NYPL Help Center*. http://catalog.nypl.org/screens/help_googlebooks_about.html. Accessed April 27, 2011.
- — —. "Google Books FAQ." *NYPL Help Center*. http://catalog.nypl.org/screens/help_googlebooks_faq.html. Accessed April 27, 2011.
- Princeton University Library. "Google Books Library Project." 2010. <http://library.princeton.edu/google/>. Accessed January 17, 2012.
- "Settlement FAQ." *SU LAIR*, Stanford University, 2008. <http://lib.stanford.edu/google-books/settlement-faq>. Accessed April 27, 2011.
- "Stanford and Google Book Search: Statement of Support and Participation." *SU LAIR*, Stanford University, 2005. <http://lib.stanford.edu/google-books/statement-support-participation-2005>. Accessed April 27, 2011.
- Stanford University Amicus Letter*. The Authors Guild, Inc., Association of American Publishers, Inc., et al. vs. Google, Inc., Case No. 05 CV 8136-DC (2009). Retrieved April 27, 2011 from: <https://lib.stanford.edu/files/GBS-StanfordAmicus-9-8-09-2.pdf>.
- Stanford University Libraries. "Stanford Google Library Project: FAQ." 2006. <https://lib.stanford.edu/google-books/faq>. Accessed January 17, 2012.
- "Stanford's Role in Google Books." *SU LAIR*, Stanford University, 2010. <http://lib.stanford.edu/google-books/stanfords-role-google-books>. Accessed April 27, 2011.
- Universiteitsbibliotheek Gent. "Google Project: In a Nutshell." 2007. <http://lib.ugent.be/info/en/project-google.shtml>. Accessed April 27, 2011.
- University of Michigan. "Michigan Digitization Project - Google Settlement Agreement FAQ." <http://www.lib.umich.edu/michigan-digitization-project/michigan-digitization-project-google-settlement-agreement-faq>. Accessed January 20, 2012.

- – – . “Settlement Agreement between Google and Plaintiffs the Authors Guild and the Association of American Publishers.”
http://www.vpcomm.umich.edu/pa/key/google_settle.html. Accessed January 20, 2012.
 - – – . “U-M University Library – Google Digitization Project: A Brief Overview.” 2004. Retrieved January 20, 2012 from:
<http://www.lib.umich.edu/files/services/mdp/overview.pdf>.
 - – – . “Google/U-M Project Opens the Way to Universal Access to Information.” December, 2004. Retrieved January 20, 2012 from:
<http://www.ns.umich.edu/index.html?Releases/2004/Dec04/library/index>.
 - – – . “UM Library/Google Digitization Partnership FAQ, August 2005.” August, 2005. Retrieved January 20, 2012 from:
<http://www.lib.umich.edu/files/services/mdp/faq.pdf>.
 - – – . “University’s Digitization Efforts.” 2011. University of Michigan Public Affairs.
<http://www.vpcomm.umich.edu/pa/key/AuthorsGuildQA.html>. Accessed September 29, 2011.
- University of Michigan, and Google, Inc. “Michigan Digitization Project - UM/Google Cooperative Agreement.” 2005. Retrieved January 20, 2012 from:
<http://www.lib.umich.edu/node/19846>.
- University of Michigan Libraries. “Michigan Digitization Project Workflow.” 2008. Retrieved January 20, 2012 from:
http://www.lib.umich.edu/files/services/mdp/MDPflowchart_v3.pdf.
- – – . “Michigan Digitization Project.” 2011. <http://www.lib.umich.edu/michigan-digitization-project>. Accessed January 17, 2012.
- University of Texas. “Statement on Use of Digital Archives.” 2007. Retrieved January 17, 2012 from: http://www.lib.utexas.edu/google/statement_of_use.html.
- University of Texas, and Google, Inc. “Cooperative Agreement.” January 8, 2007. Retrieved January 17, 2012 from:
https://www.lib.utexas.edu/sites/default/files/google/utexas_google_agreement.pdf.
- University of Texas Libraries. “FAQs About the University of Texas - Google Books Library Project.” <http://www.lib.utexas.edu/google/faqs.html>. Accessed January 17, 2012.
- University of Virginia Library. “About the U.Va./Google Book Library Project.”
<http://www2.lib.virginia.edu/press/uvagoogles/about.html>. Accessed January 17, 2012.
- – – . “Frequently Asked Questions (FAQs) About the University of Virginia/Google Book Library Project.” <http://www2.lib.virginia.edu/press/uvagoogles/faq.html>. Accessed January 17, 2012.
 - – – . “Google Book Library Project: University of Virginia Joins Leading Research Libraries in Partnership with Google to Increase Discovery of Knowledge — and to Offer Library Books to Global Audience.” November 14, 2006. Retrieved January 17, 2012 from:
<http://www2.lib.virginia.edu/press/uvagoogles/>.
- University of Wisconsin. “UW-Madison Joins Massive Google Book Project.” October 12, 2006. Retrieved January 20, 2012 from:
<http://www.library.wisc.edu/digitization/press06.html>.
- University of Wisconsin, and Google, Inc. “Cooperative Agreement.” 2006. Retrieved January

- 20, 2012 from: <http://www.library.wisc.edu/digitization/googlecontract.pdf>.
- University of Wisconsin Libraries. "University of Wisconsin - Madison Google Digitization Initiative: Principles and Values." <http://www.library.wisc.edu/digitization/#about>. Accessed January 20, 2012.
- — —. "University of Wisconsin - Madison Google Digitization Initiative: FAQs." <http://www.library.wisc.edu/digitization/faq.html#primary>. Accessed January 20, 2012.
- Van Gemert, Ed. "UW-Madison Expands Agreement with Google." July 9, 2009. Retrieved January 20, 2012 from: <http://www.library.wisc.edu/digitization/press.html>.
- Verba, Sidney. "A Letter from Sidney Verba." *Harvard University Library News Archive*, 2004. http://hul.harvard.edu/news/2004_1213_letter.html. Accessed April 27, 2011.
- Wittenborg, Karin. Letter in support of final settlement, *The Author's Guild et al v. Google*, to Denny Chin, September 3, 2009: http://www2.lib.virginia.edu/press/uvagooglesuva_librarian_letter_of_support.pdf.

Open Content Alliance

- Interviews: P1, P2, P3, P5, P6, P7, P11, P14, P15, & P18
- Biodiversity Heritage Library. "Scanning Processes." <http://biodivlib.wikispaces.com/Scanning+Process>. Accessed January 22, 2012.
- — —. "Biodiversity Heritage Library Digitization and Metadata Specifications." <https://docs.google.com/Doc?docid=0ASGG0YxTuAn2ZGd2anZ2a3pfMmZzOXZybg&hl=en&pli=1>. Accessed February 1, 2012.
- "A Call to Participate in the Open Content Alliance." <http://www.opencontentalliance.org/participate/>. Accessed February 2, 2012.
- Consortium of Academic and Research Libraries in Illinois. "CARLI Book Digitization Initiative." 2007. Retrieved January 22, 2012 from: <http://www.carli.illinois.edu/mem-serv/coll-man/09Book-Digit-Init.pdf>.
- — —. "Collections Management, OCA/Book Digitization Initiative, CARLI Book Digitization Initiative." October, 2011: <http://www.carli.illinois.edu/mem-serv/coll-man/collman-digit.html>.
- Frueh, Linda. "New Contributors to Open Content Alliance Collections." *Opencontentalliance.org* (blog). November 9, 2008. Accessed February 2, 2012. <http://www.opencontentalliance.org/2008/11/09/new-contributors-to-open-content-alliance-collections/#more-241>.
- Guren, Cliff. "Live Search Books Beta Release." *Bing Community* (blog). December 5, 2006. http://www.bing.com/community/site_blogs/b/search/archive/2006/12/05/live-search-books-beta-release.aspx.
- — —. "Live Search Books: Now with in-Copyright Content." *Bing Community* (blog). May 31, 2007. Accessed February 2, 2012. http://www.bing.com/community/site_blogs/b/search/archive/2007/06/01/live-search-books-now-with-in-copyright-content.aspx.
- Internet Archive. "Frequently Asked Questions: Texts and Books." http://archive.org/about/faqs.php#Texts_and_Books. Accessed May 1, 2012.
- — —. "Internet Archive's Terms of Use, Privacy Policy, and Copyright Policy." 2001.

- <http://archive.org/about/terms.php>. Accessed May 1, 2012.
- Internet Archive, and Boston Public Library. "The John Adams Library Collection - Cooperative Agreement." April 13, 2007. Retrieved November 15, 2011 from: <http://ia600608.us.archive.org/32/items/bostonpubliclibrary/BLPAgreementnosignatures.pdf>.
- Kahle, Brewster. "Announcing the Open Content Alliance." *Yahoo! Search Blog*, October 2, 2005. Retrieved June 8, 2008 from: <http://www.ysearchblog.com/2005/10/02/announcing-the-open-content-alliance/>.
- — —. "Bookscanning Launch and Vision of an Open Library." 2005. <http://www.archive.org/post/45685/bookscanning-launch-and-vision-of-an-open-library>. Accessed February 2, 2012.
- — —. "Open Content Alliance: Call for Participation." 2005. <http://www.archive.org/post/44311/open-content-alliance-call-for-participation>. Accessed February 2, 2012.
- — —. *Libraries Going Open*. San Francisco: Internet Archive, 2007. Retrieved January 26, 2011 from: http://www.archive.org/details/oca_test_004.
- — —. "Libraries Decide to Go Open, and 100k Books on Site ", 2007. <http://www.archive.org/post/88830/libraries-decide-to-go-open-and-100k-books-on-site>. Accessed February 2, 2012.
- — —. "Help Correct the Scans of Books." 2007. <http://www.archive.org/post/131761/help-correct-the-scans-of-books>. Accessed February 2, 2012.
- — —. "80 Libraries Going Open." 2007. <http://www.archive.org/post/159001/80-libraries-going-open>. Accessed February 2, 2012.
- — —. "A Book Grab by Google." Op-Ed. *Washington Post*, May 19, 2009. Retrieved May 19, 2009 from: <http://www.washingtonpost.com/wp-dyn/content/article/2009/05/18/AR2009051802637.html>.
- — —. "Achievements for Humanity." *Opencontentalliance.org* (blog). July 5, 2009. Accessed February 2, 2012. <http://www.opencontentalliance.org/2009/07/05/achievements-for-humanity/>.
- — —. "New Archive Book Reader." *Opencontentalliance.org* (blog). June 9, 2009. Accessed February 2, 2012. <http://www.opencontentalliance.org/2009/06/09/new-archive-book-reader/>.
- — —. "Economics of Book Digitization." *Opencontentalliance.org* (blog). March 22, 2009. Accessed February 2, 2012. <http://www.opencontentalliance.org/2009/03/22/economics-of-book-digitization/>.
- — —. "60,000 Books from Library of Congress Go Online." 2009. <http://www.archive.org/post/283352/60000-books-from-library-of-congress-go-online>. Accessed February 2, 2012.
- — —. "70,000 Books from Cornell Libraries Online Now." 2010. <http://www.archive.org/post/283343/70000-books-from-cornell-libraries-online-now>. Accessed February 2, 2012.
- — —. "Milestone Reached! 250,000 Books from Utoronto." 2010. <http://www.archive.org/post/326578/milestone-reached-250000-books-from-utoronto>. Accessed February 2, 2012.
- — —. "Digital Lending Library." 2010. <http://www.archive.org/post/312815/digital>.

- lending-library. Accessed February 2, 2012.
- — —. “In-Library Ebook Lending Program Launched.” 2011.
<http://www.archive.org/post/349420/in-library-ebook-lending-program-launched>.
 Accessed February 2, 2012.
- Kruger, Betsy. “University of Illinois at Urbana-Champaign Fy08 Large Scale Digitization Budget Proposal.” January 24, 2007. Retrieved January 22, 2012 from:
http://www.library.illinois.edu/committee/exec/supplement/2006-2007/Large_Scale_Digitization_Working_Group_FY08_Budget_Request.pdf.
- Library of Congress. “Federal Scanning Center at the Library of Congress (Fedscan): Frequently Asked Questions.” October 29, 2008. Retrieved January 22, 2012 from:
http://www.loc.gov/flicc/contracts/pdfs/FEDSCAN_IAFAQs.pdf.
- “Milestone Achieved.” *Opencontentalliance.org* (blog). December 20, 2006. Accessed February 2, 2012. <http://www.opencontentalliance.org/2006/12/20/milestone-achieved/>.
- MSN. “MSN Search Announces MSN Book Search.” October 25, 2005. Retrieved January 17, 2012 from: <http://www.microsoft.com/presspass/press/2005/oct05/10-25MSNBookSearchPR.msp>.
- Nadella, Satya. “Book Search Winding Down.” *Bing Community* (blog). May 23, 2008.
<http://blogs.msdn.com/livesearch/archive/2006/12/05/live-search-books-beta-release.aspx>.
- Open Content Alliance. “About.” <http://www.opencontentalliance.org/about/>. Accessed January 22, 2012.
- — —. “FAQ.” <http://www.opencontentalliance.org/faq/>. Accessed January 22, 2012.
- “Sloan Foundation Grant Awarded.” *Opencontentalliance.org* (blog). December 19, 2006.
 Accessed February 2, 2012. <http://www.opencontentalliance.org/2006/12/19/sloan-foundation-grant-awarded/>.
- “The Internet Archive Receives Grant from Alfred P. Sloan Foundation to Digitize and Provide Open Online Access to Historical Collections from Five Major Libraries.” December 20, 2006. Retrieved February 2, 2012 from:
<http://www.archive.org/serve/uploaded/brewster-OCAreleaseFINAL-1.doc>.
- Tufts University. “Books for All: Library Consortium to Offer Free Online Access to 34 Million Volumes.” 2007.
<http://tuftsjournal.tufts.edu/archive/2007/october/briefs/books.shtml>. Accessed January 22, 2012.
- University of California Libraries. “UC Libraries Partner with Technology Companies and Non-Profits to Provide Free Access to Digitized Books.” October 3, 2005. Retrieved January 22, 2012 from: <http://www.universityofcalifornia.edu/news/article/20320>.
- Yahoo. “Global Consortium Forms Open Content Alliance to Bring Additional Content Online and Make It Searchable.” October 3, 2005. Retrieved January 17, 2012 from:
<http://pressroom.yahoo.net/pr/ycorp/176055.aspx>.

Appendix C: Interview Protocol (Generic)

Interview Protocol for [INSTITUTION]

Opening

- Personal introductions/small talk
- Explain the **purpose of the study**:
 - To understand how [Google's/ the OCA's] effort to digitize all the books in the world came to be defined and designed in the way that it has, and the roles played by different kinds of individuals in shaping those definitions and designs. I'll be doing a similar analysis of the [Open Content Alliance/Google Books], as well as the early years of the Boston Public Library and the Carnegie Library program in the late nineteenth century, and comparing the different approaches taken to broadening access to knowledge in those four projects.
- Explain the **interview structure**: Three parts:
 - We'll start off with some basic demographics and information about you;
 - The second section will ask you to jog your memory about the history of the project and your involvement with it,
 - The third section will go a little deeper into certain areas – specifically, how you and others on this project think about concepts like comprehensive coverage and universal access, and what the network of people responsible for shaping this project might look like.
 - The second and third sections might overlap a little bit, but that's ok.
- **Check whether audio is OK** (even if they've already said OK by email)
 - If ok, turn on recorder and double-check that it's working
 - SAY THE PARTICIPANT NUMBER AND DATE INTO THE TAPE
 - If not ok to record, explain that you will take notes
- Explain **process**:
 - The interview is confidential; I won't reveal individual interview details; all interview records will be assigned a code and any other identifiers will be removed.
 - If I ask any questions that you either cannot or would prefer not to answer, just tell me so, and we'll move on.
 - After the interview, I'll transcribe the tape (/notes) and send a copy to you for confirmation (at which point you can request changes or deletions); after transcription, the tape will be erased.
- **ASK: Do you have any questions before we get started?**

Demographics/personal History:

- Age group: 20-29 30-39 40-49 50-59 over 60
- Gender: Male Female Other
- Where are you from, originally?

- Could you briefly describe your educational background?
- Could you briefly describe your professional background? (e.g. are they an engineer? A librarian?)
- Where do you work now?
 - What's your role?
 - How long in that role?

Personal involvement with [Google Books/OCA]

- How did you come to be involved with the [Google Books/OCA] project? (When?)
 - What is/was your role?
 - How long in that role?
 - What aspects of the project were you/have you been especially involved with?
- What have been the major milestones in the project, from your perspective?
 - How were you involved in those milestones, if you were?
- Potential follow-ups:
 - Are there particular pieces of the project, or how it has evolved, that you feel especially responsible for (or proud of)?
 - Is there anything about the project that you wish had happened differently?
 - As you know, there are other competing projects with similar goals; what do you think makes Google Books distinctive or more likely to succeed?

- (If they have left:) What made you decide to leave the project? How long were you involved?

General impressions of [Google Books/OCA]

- Why do you think [your institution], [Google/IA], and the other founding organizations embarked on this project? What do you think that those making the decision hoped to achieve?
 - Do you think it is living up to those expectations / hopes? (Do you think it will (or could)?)
- Probe the issue of universal coverage
 - Various people have said that the ultimate goal of Google Books is to make every book in the world searchable – do you think that kind of comprehensiveness is possible, in terms of collections? (Can you really get **every** book?) [For OCA: Brewster Kahle has frequently made comments to the effect that his goal is universal access to all knowledge – do you think that kind of universality is possible, in terms of digital book collections?]
 - Why/why not?
 - Do you think this project is likely to achieve it?
 - Are there any kinds of things this project seems to be systematically missing (for practical reasons, legal reasons, etc.)?

- Is there any kind of order to what gets scanned? Have particular kinds of things been scanned or posted before others? What kinds of concerns seem to set the order?

- Probe the issue of universal access
 - What kind of user (or users) do you personally have in mind when thinking about this project? Who do you see using this resource? (Who do you picture sitting in front of the screen at the other end?)

 - Do you think that those designing the access systems – interfaces, information architecture, etc. – have the same kind of user in mind? Are there multiple types?
 - Where do you think these ideas about the users come from?

 - Do you know of any formal strategies used in the design processes for the interfaces or other access systems that might tend to orient the system design toward particular sorts of users? (E.g., are there personas used, any kind of systematic user testing, more ad hoc?)

- Social network assessment
 - Who do you think has played an especially strong role in shaping the project? How so?
 - Are there groups or individuals whom you wish had had a stronger role in shaping the project?
- Snowball sampling
 - Is there anyone you think it would be especially important for me to interview for this study?
 - Explain: I'll follow up with them via email; I won't reveal who recommended them for the study unless you give me permission to do so, or if you introduce me.
 - Would you be comfortable providing contact information, and/or making an introduction via email?
- Are there any other questions you think I should have asked you, or should ask others?

Appendix D: Contact Summary Form

Participant # _____

Date _____

Contact Summary Form

Contact type:

In-person _____

Skype _____

Other _____

Case: _____

Site: _____

Duration: _____

1. What were the main issues or themes that struck you with this contact?

2. Summarize the information you got (or failed to get) on each of the target questions:

	<u>Question</u>	<u>Information</u>
	Their story	
Personal Involvement	Major milestones	
	Proud moments	
	Wish were different	
	What distinguishes their project	
	Why they left	
	Why do this	
General Impressions	Things missed	
	Things given priority	
	Their imagined user	
	The project's imagined user	

Participant # _____
Date _____

Formal UCD
strategies

How many people

Most influential
ppl

Wish were more
influential

Catchall/Suggested
Q's

3. Anything else that struck you as salient, interesting, illuminating, or important in this contact?

4. What new (or remaining) target questions do you have for interviews with others on this project (or the other project)?

Age

20-29 ____
30-39 ____
40-49 ____
50-59 ____
60+ ____

Gender

Male ____
Female ____

Where are you from, originally? _____

Educational background

No college ____
Some college ____ at _____
Bachelor's Degree ____ from _____
Master's Degree ____ from _____
PhD ____ from _____

Professional Background

Appendix E: Google Books Library Project Timeline

2002

- Spring: Larry Page first introduces the idea of digitizing the University of Michigan Library on a visit to campus (P8, P10, "MDP Timeline")
- November 2002: Talks get underway in earnest between Google and UM librarians, administrators, and lawyers; Larry Page still personally involved (P10, "MDP Timeline")

2003

- March 3-4: Team from Google visits Michigan and begins technical planning ("MDP Timeline")
- Spring-Summer: Monthly calls between Google & UM (P9, "MDP Timeline")
- Oxford, NYPL, Harvard brought in on discussions (P13, P18, Carr 2005b)
- December: Google Print publisher program first publicly discussed ("Google Print" 2003)
- Late 2003 – Early 2004: Library negotiations on file specifications; technology development at Google ("MDP Timeline")

2004

- April 19: University of Michigan contract signed; specified that the first six months would constitute a “pilot” during which either party may withdraw for any reason (University of Michigan and Google 2004, "MDP Timeline")
- June-October: Testing of technology and output with Google at UM; role of library partnerships manager(s) created (P12, Last 2007, "MDP Timeline")
- October 5: First publisher project books go live at print.google.com (Milliot and Zeitchik 2004); “Initial partners include Penguin, Hyperion, Scholastic, Houghton Mifflin and a number of university presses. Notable resistance comes from Random House” (Albanese 2009b)
- October: Production scanning begins at University of Michigan ("MDP Timeline")
- November: Google Scholar announced (Banks 2005)
- Fall: Other four G5 contracts signed (Stanford, Harvard, Oxford, NYPL) (P8)
- December 14: Google Print Library Project officially announced (Harvard internally 1 day before everyone else) (Google 2004, Markoff and Wyatt 2004, Quint 2004, Verba 2004)

2005

- Scanning begins at G5 libraries; Google starts building out additional scanning centers (P4, P16)

- Dan Clancy hired to manage project (P4)
- May: first library project books go live on Google (Last 2007)
- June 17: UM contract made available through Google-Watch.org and UM website following FOIA request by Google-Watch (Helm 2005b)
- August 11: Google announces hiatus on scanning in-copyright books (Smith 2005)
- September 20: AG lawsuit filed (Helm 2005a)
- October 19: AAP lawsuit filed (Helm 2005a)
- November 1: Google resumes scanning copyrighted works from libraries, after discussions with publishers fall apart. (Helm 2005a)
- November (17?): Name change (Google Print -> Google Book Search) (Gohring 2005)

2006

- Google hits 1 million volumes; rate of scanning increases; Doug Kuch hired to run logistics for GBLP (P4, P12).
- Three US scanning centers in operation (to my knowledge): Mountain View (CA), Ann Arbor (MI), Cambridge (MA). Also one scanning center in the UK, at Oxford. (P8, P12, P16, P17)
- Four partners added: University of California (August), Universidad Complutense de Madrid (September), University of Wisconsin (October), University of Virginia (November) (Colvin 2006, University of Virginia Library 2006, University of Wisconsin 2006, Crawford 2007).
- January: Google announces that it has formed a “fact-checking brigade” to clear up misconceptions about its digitization project (Foster 2006).
- June 6: Copyright lawsuit filed by La Martinière Groupe, joined June 10 by French Publishers’ Association (Sayer 2006).
- June 29: German lawsuit (filed by publisher WBG) thrown out because Google was not actually scanning WBG’s books; they had opted out (Wray 2006).
- August 31: MBooks announced (Albanese 2006a, Quint 2006b, Crawford 2007).
- September 5: Google starts to allow PDF downloads of public domain works; also starts to include “Find in a library” links “to WorldCat.org and international library catalogs” (Albanese 2006a, Quint 2006b).
- September 26: Universidad Complutense de Madrid joins project (Crawford 2007).
- October: University of Wisconsin joins project (Crawford 2007).
- November 14: University of Virginia joins project (Crawford 2007).

2007

- Twenty partners added: University of Texas (January), Biblioteca de Catalunya (January), Princeton University (February), Bavarian State Library (March), Lausanne

University (May), Ghent University Library (May), University of Mysore (May), the Committee on Institutional Cooperation (12 universities, of which two were already participating; June), Keio (July), Cornell University (August), Columbia University (December) (Grogg and Ashmore 2007).

- NYPL begins to offer access to its scans via its catalog (New York Public Library)
- February 5: *The New Yorker's* Jeffrey Toobin becomes the first journalist to predict a settlement in the AAP/ AG lawsuit (Toobin 2007).
- April 2: Stanford launches its Copyright Renewal Database (Calter 2007).
- November: Amazon announces the Kindle (Last 2007).
- By the end of the year: Google has signed up 10,000 publisher partners (Albanese 2009b).

2008

- Google finishes scanning materials from Harvard and Oxford (I suspect NYPL also finished around this time, but have found no direct substantiation for that) (P13, P16).
- February: University of Michigan reaches 1 million books scanned (MLibrary 2008)
- June 24: Google releases a downloadable XML file “containing U.S. copyright renewal records for books published from 1923 to 1963.” (Quint 2008)
- October 13: HathiTrust launched (Wilkin 2008)
- October 28: Settlement agreement announced in AAP/ AG lawsuit (Albanese 2008a)
- By October: Google reaches 7 million volumes (Leetaru 2008)

2009

- February: Darnton publishes his first NYRB piece, which crystallizes opposition to the settlement by “raising concerns about a Google monopoly over book content” (Albanese 2009b)
- March: Publisher partner program has more than 20,000 members, including the initially opposed Random House – and double the number it had two years ago (Albanese 2009b)
- April: A group of authors led by Gail Knight Steinbeck successfully lobbies for an extension on the settlement’s opt-out deadline for publishers and authors from May 5 to September 4, giving all involved (or simply interested) an additional four months to examine the details of the agreement. [May have been its initial death knell...] (Albanese 2009c)
- September 10: Congressional hearing on GBS Settlement; strong critiques from Register of Copyrights Marybeth Peters (Albanese 2009d)
- October 7: Settlement fairness hearing (Albanese 2009a)

- November 13: Revised Settlement Agreement submitted to the court. New provisions include exclusion of all international works except those from UK, Australia, and Canada, as well as shift in revenue policy for orphan works (instead of diverting those profits to authors who have registered, they would go toward the costs of locating absent rightsholders; after 10 years, they would be dispersed to literacy-based charities in US, UK, Canada, and Australia). (Jones 2009) Some library contracts are revised in its wake (in anticipation of its approval) (University of Michigan and Google 2009, University of Texas and Google 2009, University of Wisconsin and Google 2009)
- December: French court loss; no more scanning in-copyright books in France

2010

- Partners added: Italian Ministry of Culture (March), Austrian National Library (June), Dutch National Library (July)
- February 18: Revised settlement fairness hearing (big slams from DoJ) (Albanese 2010b)
- April 7: Lawsuit filed by artists and photographers against Google over Book Search, having been denied the opportunity to join the AAP/AG settlement class (Albanese 2010a)
- By July: Google has reached 12 million volumes (Orwant 2010b)
- December 6: Google launches Google eBookstore, starts selling eBooks (Albanese 2010d)

2011

- Partners added: Czech National Library (February), British Library (June)
- March 22: Settlement rejected on copyright and antitrust grounds (*Rejected Settlement Opinion* 2011, Albanese 2011b)
- As of August: Publisher program has over 30,000 partners (Albanese 2011a)
- October: Google shuts down Mountain View scanning center, leaving Ann Arbor as the sole remaining scanning center in the United States; rate of scanning decreased (P8, P17)

2012

- January: HathiTrust reaches 10 million volumes (HathiTrust Digital Library 2012)
- October: Google and the Association of American Publishers reach a settlement agreement; most of its exact terms remain private (Sporkin 2012b).
- Google reaches 20 million volumes (Howard 2012).

Appendix F: Open Content Alliance Timeline

1980 - mid-1990s

- Brewster Kahle begins to think about building a universal digital library, but recognizes several gaps in the existing technology, and sets about trying to fill them, working on increasing the capacity of digital storage, building out digital networks, and working out online revenue streams for publishers (Kahle 2011).

1995

- Kahle sells WAIS, an early publishing and distributed search system, to AOL for \$15 million (Hogge 2005, "Staff Bios").

1996

- Kahle founds web analytics company Alexa Internet, named for the Library of Alexandria, in collaboration with Bruce Gilliat. As part of its functionality, Alexa crawls and archives webpages (Tong 2002, "Staff Bios").
- Kahle founds the Internet Archive, which uses Alexa's archived web-crawls as the foundation for its archive of internet content, made accessible via the Wayback Machine (Hogge 2005).

1999

- Kahle sells Alexa Internet to Amazon for \$250 million; the Internet Archive retains the right to receive web crawls from Alexa (Hogge 2005, Hardy 2009, "Technology").

2000

- The Million Book Project, an NSF-funded large-scale digitization project geared toward addressing a particular set of technological research questions, is initiated by a group of computer scientists at Carnegie Mellon, in collaboration with universities in China and India, the Biblioteca Alexandrina in Egypt, and other partners (St. Clair 2008, 152-53).

2004

- The Internet Archive signs on to assist the Million Book Project (by then also referred to as the Universal Library Project) with permanent archiving, quality control, and materials acquisition ("Frequently Asked Questions" 2007). Following on their experiences with Million Book, IA begins to develop its own book scanning technology (both hardware and software), which would eventually be called the Scribe (Kahle 2011).
- Autumn: The Internet Archive and the University of Toronto begin a pilot project to test the IA's scanning process. Within the next year, they scan about 2,000 books (Carlson and Young 2005, Bengtson 2006).

- December 14: Google announces the Google Books Library Project (Google 2004).

2005

- Early 2005: Conversations begin within the Internet Archive and its social sphere about starting up a more open book scanning project, as an alternative to Google's; initial thought was to fund the project via contributions from libraries (P1).
- Early/Mid 2005: Sumir Meghani of Yahoo! approaches the Internet Archive, proposes the concept of the Open Content Alliance, a collaborative project to scan works out of copyright or otherwise openly available for scanning. Yahoo! puts up \$150,000 in funding to start, targeting a collection of works in American Studies at the University of California (P1, P3, Albanese 2005a, Johnson 2007, 5).
- October 2: Open Content Alliance announced by Brewster Kahle on the Yahoo! Search Blog (2005a). Founding partners include:
 - The Internet Archive (technology/content/hosting/administration)
 - Yahoo! (technology/indexing/financial)
 - Adobe Systems (technology)
 - The European Archive (content)
 - HP Labs (technology)
 - The UK National Archives (content)
 - O'Reilly Media (content)
 - Prelinger Archives (content)
 - University of California (content)
 - University of Toronto ("Consortium Forms OCA" 2005)
- By October 14: The project is endorsed by the Association of Learned and Professional Society Publishers (ALPSP) (Carlson and Young 2005).
- October: Partners added: Lulu (a print on demand service), LibriVox (a producer of audio editions), the Biodiversity Heritage Library, the Smithsonian Institution Libraries, and eleven university libraries (McMaster, Memorial University of Newfoundland, the University of Ottawa, University of British Columbia, York University, Columbia University, Emory University, Johns Hopkins University, the University of Virginia, Rice University and the University of Pittsburgh) (Auchard 2005, Crawford 2005, Suber 2005).
- October 25: Microsoft joins the OCA, pledging funding and tech/scanning assistance (Auchard 2005); simultaneously announces MSN Book Search ("MSN Search Announces MSN Book Search" 2005, Albanese 2005a, Crawford 2005).
- October 27: The Research Libraries Group (RLG) signs on to provide the OCA with bibliographic information from its union catalog ("RLG Joins Open Content Alliance" 2005, Crawford 2005).

- By November 2: The project is endorsed by the Association of American Publishers (AAP) and the American Association of University Presses (AAUP) (Suber 2005).
- November 4: British Library signs on to scan 100,000 books with Microsoft ("Microsoft Scans British Library" 2005, Crawford 2005, Kupferschmid 2005).
- November 8: Official launch of OCA book scanning; Microsoft pledges to fund the digitization of 150,000 by the end of 2006 (Kahle 2005b).
- November 9: WSJ reports that the IA/Toronto pilot project has scanned 2800 books, at a cost of \$108,250, over the course of the past year (Crawford 2005).
- December 29: a group of 27 Canadian research libraries announces that they are jointly establishing a project called Alouette Canada, a digitization alliance intended to work collaboratively with the OCA (Crawford 2006).

2006

- January: Open Library (www.openlibrary.org) first goes live as a demonstration site (Notess 2006).
- First OCA scanning center opens at the University of Toronto (P5).
- March: Partner added: University of North Carolina-Chapel Hill (Library and SILS; providing content and expertise, respectively) ("UNC-Chapel Hill Library and Library School Join Open Content Alliance" 2006, Albanese 2006c).
- September: The University of California joins Google Book Search; Kahle sees this as a betrayal of the OCA's principles on UC's part, going so far as to claim that UC is "effectively giving their library to a single corporation," despite its then-ongoing participation in the OCA (P1, P2, Albanese 2006b).
- December 6: Microsoft Live Search Books releases its first beta version (at <http://books.live.com> - now defunct) (Guren 2006).
- December 19: IA receives \$1 million from the Sloan Foundation to scan specific collections from the Boston Public Library, The Getty Research Institute, The Metropolitan Museum of Art, UC-Berkeley's Bancroft Library, and Johns Hopkins ("Sloan Foundation Grant Awarded" 2006, "Internet Archive Receives Grant" 2006, Internet Archive and Boston Public Library 2007). Three out of five of these institutions are new partners for the OCA.
- December 20: IA announces that it has digitized and made available 100,000 books, largely from members of the Open Content Alliance ("Milestone Achieved").

2007

- March: The OCA reaches 130,000 volumes scanned, all available via the IA Text Archive (Notess 2007).

- April: Partner added: University of Illinois ("U of Illinois Joins Open Content Alliance" 2007).
- June 25: The Internet Archive successfully petitions to be declared a library by the State of California, in order to gain eligibility for state-administered federal grants (Kahle 2007a, A. McCoy 2007).
- July 16: At Kahle's request, open access advocate and entrepreneur Aaron Swartz signs on to help build the architecture for Open Library. His goal in the design, as he puts it, is to create "a website with a page for every book, collecting everything we can find out about it from libraries, publishers, reviewers, and of course, book lovers" (Swartz 2007, Kniffel 2008, quoting Swartz).
- October: The OCA reaches 200,000 volumes scanned. Eight scanning centers are in operation, in three countries: the US, Canada, and the UK (Goth 2007, Kahle 2007b, Ashmore and Grogg 2008).
- October 25: IA releases a rewritten version of its scanning software, Scribe2, which promises greater format flexibility, less bandwidth usage, and support for new cameras (Internet Archive 2007, Kahle 2007b).
- November: Partner added: Boston Library Consortium (19 member libraries, all contributing public domain materials only, self-funded). The BLC publicly announces that it was approached by Google first, but rejected them in favor of OCA ("Boston Library Consortium and Open Content Alliance to Provide Digitized Books" 2007, Albanese 2007a, Hane 2007).
- November 15: A set of OCA partners – IA, BPL, MBL-WHOI, and Universidad Francisco Marroquín – announce a plan to scan out-of-print, in-copyright works for distribution via a new form of digital interlibrary loan, which they will develop (Albanese 2007c, Kahle 2007b).
- December: The OCA reaches 250,000 volumes scanned (Hane 2007).
- December 15: Yale University signs on with Microsoft to scan up to 100,000 books *outside* of the OCA, and on terms more like Google's than like IA's (Albanese 2007e).

2008

- January-April: With funding from Microsoft, IA deploys five more US-based scanning centers under the banner of the OCA (P5, Kahle 2008).
- February: The Boston Public Library begins to offer scan-on-demand interlibrary loan (ILL) services for public domain works using its onsite Scribe workstations. This reduces the turnaround time from weeks to days, and makes it possible to fulfill ILL requests that would otherwise have been denied due to the condition and/or rarity of the item requested (Colford 2008).

- February 19: Partner added: Triangle Research Libraries Network, a consortium composed of the research libraries at Duke University, North Carolina Central University, North Carolina State University, and the University of North Carolina at Chapel Hill (the last of which was already an OCA partner) ("Triangle Research Libraries Network" 2008, "TRLN Libraries" 2008, "TRLN Member Libraries" 2008).
- May 23: Microsoft announces it is bowing out of book scanning, having spent \$10 million on its efforts. The company's departure leaves a major gap in OCA funding (P5, Albanese 2008d, Guess 2008, Kahle 2008, Nadella 2008).
- August: Maura Marx, then head of the Boston Public Library's Digital Content Program, is hired to be the first Executive Director of the OCA. However, she never actually assumes the role, but instead founds a separate initiative, Open Knowledge Commons, with Sloan funding (P1, P11, "People" 2008, Berry 2009).
- November: HathiTrust launched, incorporating scanned content from OCA as well as Google Books and other digitization projects (Albanese 2008b).
- December: Open Library and Boston Public Library jointly begin to offer a scan-on-demand service for public domain works that have been indexed by Open Library, but have not yet been made available in full text ("Have a Hand in Scan-on-Demand" 2008).

2009

- January: The OCA reaches 1 million volumes scanned – including 300,000 donated by Microsoft after the discontinuation of Live Search Books (O'Leary 2009).
- July 5: Last mention of the term "Open Content Alliance" on the organization's own blog (Kahle 2009b). After about this point, the project under that name is effectively defunct, though various pieces of it persist, and the term still pops up occasionally in discussions of book digitization.
- December: IA and the OCA form the Open Book Alliance to oppose the Google Books Settlement Agreement. Yahoo, Microsoft, Amazon, the Special Libraries Association and the New York Library Association soon join (Oder, et al. 2009).

2010

- February: The IA debuts BookServer, "a distributed system for lending and vending on the Internet" at the O'Reilly Media Tools of Change for Publishing Conference. It allows individuals to buy or check out in-copyright but out-of-print materials, dovetailing with Open Library and connecting with libraries and retailers (Hadro 2010a).
- Spring: IA works with the City of San Francisco to hire over 125 workers for its scanning project, as subsidized labor under the Temporary Assistance for Needy Families (TANF) program (P5, Miller 2010).
- May 6: In the final post on the Open Content Alliance blog, Kahle announces that the IA will be making 1 million books, both in and out of copyright, accessible to the print

disabled via Open Library in the open DAISY talking book format (Hadro 2010b, Kahle 2010d).

- June: The first 200 or so ebook versions of out-of-print, in-copyright books go live for lending via Open Library. They are readable for two-week periods using Adobe Digital Editions Software (Rapp 2010).²⁶¹

2011

- April: IA announces that 85,000 in-copyright, out-of-print titles, contributed by 150 public and academic libraries, will be made available via Open Library, but only to patrons actually physically located in those 150 libraries (though once patrons download the books, they can use them on their personal devices outside the library for the duration of the loan) (Rapp 2011a).
- June: OCLC researchers develop “oclcBot,” a piece of software that matches up records from Open Library to records from OCLC, checks to see if the Open Library has an OCLC number (a unique identifier commonly used across library systems), and inserts one if none is present (Rapp 2011c).
- July: Kahle announces the establishment of the Internet Archive’s physical book archive, hoping to obtain “one copy of everything ever published.” The archive is launched with an initial collection of 450,000 items, accumulated as part of IA’s various digitization efforts, and seeks to build its collection through donations and by gathering up items deaccessioned by other libraries (P5, Rapp 2011b).
- October: The state librarians of all 50 U.S. states vote unanimously to enter into a memorandum of understanding with IA, pledging their support for the Open Library’s online lending program (Kelley 2011).

²⁶¹ At the time of this writing, IA continues to digitally lend in-copyright books – and oddly enough, nobody has sued them for it yet, although some have admonished them for “stealing” (e.g., Lyons 2011).

Works Cited

- A&G Total Media. 2007. *Libraries Going Open!* San Francisco: Internet Archive. Web Video. Accessed via: http://archive.org/details/oca_2007_movie
- "Act of Incorporation." 1911. In *Charter of Carnegie Corporation of New York*. New York: Carnegie Corporation of New York.
- Adam, Nabil R., and Shamim Naqvi. 1996. "Universal Access in Digital Libraries." *ACM Computing Surveys* 28, no. 4es (December).
- Adam, Nabil, and Yelena Yesha. 1996. "Strategic Directions in Electronic Commerce and Digital Libraries: Towards a Digital Agora." *ACM Computing Surveys* 28, no. 4: 818-35.
- Adler, Richard, and Melissa Levine. 2013. "Rights, Research, Results: The Copyright Review Management System Projects." Presentation given at the CNI Membership Meeting. San Antonio, TX. April 5.
- Akrich, Madeleine. 1992. "The De-Description of Technological Objects." In *Shaping Technology/Building Society: Studies in Sociotechnical Change*, edited by Wiebe Bijker and John Law. 205-24. Cambridge, MA: MIT Press.
- Albanese, Andrew. 2005a. "AAP Sues Google over Scan Plan." *Library Journal* 130, no. 19 (November 15): 17-18.
- — —. 2005b. "Google Pauses Scan Plan." *Library Journal* 130, no. 14 (September 1): 16.
- — —. 2006a. "Google Scans, Contract Released." *Library Journal* 131, no. 16 (October 1): 14-15.
- — —. 2006b. "UC Joins Google's Scan Plan." *Library Journal* 131, no. 14 (September 1): 14-15.
- — —. 2006c. "UNC Library, SILS Join Content Alliance." *Library Journal* 131, no. 6 (April 1): 21-22.
- — —. 2007a. "BLC, OCA Join in Digitization Effort." *Library Journal* 132, no. 17 (October 15): 15-16.
- — —. 2007b. "Boston Library Consortium Partners with OCA on Digitization Project." *Library Journal* (September 28).
- — —. 2007c. "OCA to Scan Orphan Works." *Library Journal* 132, no. 19 (November 15): 16-17.
- — —. 2007d. "Scan This Book!" *Library Journal* 132, no. 13 (August 15): 32-35.
- — —. 2007e. "Yale, Microsoft Join in Scan Plan." *Library Journal* 132, no. 20 (December 15): 19.
- — —. 2008a. "Google Book Deal Gets 'Preliminary' Court Approval; More Boosters Emerge." *Library Journal* (21 November).
- — —. 2008b. "HathiTrust Is Launched." *Library Journal* 133, no. 18 (November 1): 13.
- — —. 2008c. "Libraries Reserve Early Comment, but Some See Bright Side in Google Book Search Settlement." *Library Journal* (October 28).
- — —. 2008d. "Microsoft Gives up Scan Plan." *Library Journal* 133, no. 12 (July 15): 14-n/a.
- — —. 2008e. "One for All? As Google Deal Is Evaluated, Critics Question Single Library Terminal." *Library Journal* (November 11).
- — —. 2009a. "Court Acknowledges More Than 400 Submissions in Google Settlement." *Publishers Weekly* (16 September).
- — —. 2009b. "Deal or No Deal." *Publishers Weekly* 256, no. 21 (May 25): 23-25.
- — —. 2009c. "Extension, Possible Review Challenge Google Settlement." *Publishers Weekly*, 6.
- — —. 2009d. "U.S. Register of Copyrights Slams Google Book Search Settlement." *Publishers Weekly* (10 September).
- — —. 2010a. "Artists and Photographers Sue over Google Book Search." *Publishers Weekly* (7 April).
- — —. 2010b. "At Google Fairness Hearing, DoJ Justice Slams Settlement." *Publishers Weekly* (18 February).

- — —. 2010c. "Google Launches Google Ebooks, Formerly Google Editions." *Publishers Weekly* (December 6).
- — —. 2010d. "Google Shakes up the E-Book Market." *Publishers Weekly* (December 13).
- — —. 2011a. "The Breakup: Are the Google Settlement Parties Headed for Splitsville?" *Publishers Weekly* 258, no. 31 (August 1): 18.
- — —. 2011b. "What Comes Next? With the Google Settlement Rejected, Parties Ponder Future." *Publishers Weekly* 258, no. 13 (March 28): 3.
- — —. 2012. "Appeal Filings Outline Authors Guild's Objections to HathiTrust Opinion." *Publishers Weekly*, December 3.
- American Library Association. 2008. "ALA Library Fact Sheet 22 - the Nation's Largest Libraries." ALA.
http://www.ala.org/ala/professionalresources/libfactsheets/ALA_print_layout_1_44591_44591.cfm. Accessed February 3, 2011.
- American Library Association, Association of College and Research Libraries, and Association of Research Libraries. 2009. "Library Association Comments on the Proposed Settlement." In *The Authors Guild, Inc., Association of American Publishers, Inc., et al. vs. Google, Inc.*, Case No. 05 CV 8136-DC.
- Anderson, Florence. 1963. *Carnegie Corporation Library Program, 1911-1961*. New York: Carnegie Corporation of New York.
- Anderson, Nate. 2007. "Google to Scan 800,000 Manuscripts, Books from Indian University." *Ars Technica*. <http://arstechnica.com/uncategorized/2007/05/google-to-scan-800000-manuscripts-books-from-indian-university/>. Accessed July 8, 2013.
- "Andrew Carnegie on the Negro's Progress." 1910. *New York Times*, February 6.
- Ann Arbor District Library. 2008. "Ann Arbor District Library Rules of Behavior." <http://www.aadl.org/aboutus/policies/behavior>. Accessed August 2, 2012.
- Anonymous Syracuse Student. 2009. "The New York Public Library/Google Books Partnership." *IST 677, Spring 2009* (blog). March 24. <http://ist677-2009.blogspot.com/2009/03/new-york-public-librarygoogle-books.html>. Accessed July 24, 2013.
- Armstrong, Kim. 2011. "One Million Books Scanned and Returned to CIC University Libraries." *Inside Google Books* (blog). February 4. <http://booksearch.blogspot.com/2011/02/one-million-books-scanned-and-returned.html>. Accessed March 31, 2011.
- Ashmore, Beth, and Jill E. Grogg. 2008. "The Race to the Shelf Continues - the Open Content Alliance." *Searcher* 16, no. 1 (January): 18-23.
- Auchard, Eric. 2005. "Microsoft Joins Yahoo on Digital Library Alliance." *Yahoo News*, October 26.
- Badger, Brandon. 2009. "Download over a Million Public Domain Books from Google Books in the Open Epub Format." *Inside Google Books* (blog). August 26.
<http://booksearch.blogspot.com/2009/08/download-over-million-public-domain.html>. Accessed July 8, 2013.
- Bailey, Charles W., Jr. 2010. "Google Books Bibliography." <http://www.digital-scholarship.org/gbsb/gbsb.htm>. Accessed May 20, 2010.
- Baksik, Corinna. 2006. "Fair Use or Exploitation? The Google Book Search Controversy." *portal: Libraries & the Academy* 6, no. 4: 399-415.
- Band, Jonathan. 2004. "The Google Library Project: The Copyright Debate." OITP Technology Policy Brief. Washington, DC: American Library Association. January.

- — —. 2008. "A Guide for the Perplexed: Libraries & the Google Library Project Settlement." Washington, D.C.: American Library Association and the Association of Research Libraries.
- — —. 2010. "The Long and Winding Road to the Google Books Settlement." *John Marshall Review of Intellectual Property Law* 9, no. 2: 226-329.
- — —. 2011. "A Guide for the Perplexed Part IV: The Rejection of the Google Books Settlement." Washington, DC: American Library Association, Association of College and Research Libraries, and Association of Research Libraries.
- Banerji, Pratip. 2011. "The First Google Ebooks-Integrated E-Reader: Iriver Story Hd." *Inside Google Books* (blog). July 11. <http://booksearch.blogspot.com/2011/07/first-google-ebooks-integrated-e-reader.html>. Accessed July 8, 2013.
- Banks, Marcus A. 2005. "The Excitement of Google Scholar, the Worry of Google Print." *Biomedical Digital Libraries* 2: 1-3.
- Barnett, Doug, and Travis Willmann. 2007. "The University of Texas Libraries Partner with Google to Digitize Books." January 19.
- Barrett, Brian. 2011. "Google's Secret Class System." *Gizmodo* (blog). April 29. <http://gizmodo.com/5797022/googles-secret-class-system>. Accessed July 18, 2013.
- Bates, Joshua. 1852. Letter to Thomas W. Ward, October 1. Courtesy Boston Public Library: Rare Books & Manuscripts.
- Bayerische StaatsBibliothek. n.d. "Mass Digitisation within the Framework of a Public-Private Partnership between the Bayerische Staatsbibliothek and Google." <http://www.bsb-muenchen.de/Mass-Digitisation-within-the-F.1842+M57d0acf4f16.0.html>. Accessed July 8, 2013.
- Belli, Robert F. 1998. "The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys." *Memory* 6, no. 4: 383-406.
- Bengtson, Jonathan B. 2006. "The Birth of the Universal Library." *Library Journal* 131, no. 6 (Spring): 2-7.
- Benson, Heidi. 2005. "A Man's Vision: World Library Online - Brewster Kahle Hopes to Realize His 25-Year Dream of an International Book Archive." *San Francisco Chronicle*, A1.
- Berelson, Bernard. 1971. *Content Analysis in Communication Research*. New York: Hafner.
- Berger, Peter L., and Thomas Luckmann. 1966. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Anchor Books.
- Berry, John N., III. 2009. "Chicago Hope." *Library Journal* 134, no. 10 (June 1): 22.
- Bertram, James. 1911. "Notes on Library Bilding." New York: Carnegie Corporation of New York.
- — —. 1915. "Notes on the Erection of Library Bildings." In *Free to All: Carnegie Libraries & American Culture, 1890-1920*, edited by Abigail Van Slyck. Chicago: University of Chicago Press.
- — —. 1916. "Notes on the Erection of Library Bildings." New York: Carnegie Corporation of New York.
- Biblioteca Complutense. n.d.-a. "Búsqueda De Libros Electrónicos UCM-Google." <http://www.ucm.es/BUCM/atencion/25403.php>. Accessed April 27, 2011.
- — —. n.d.-b. "Proyecto De Digitalización Biblioteca Complutense-Google." <http://www.ucm.es/BUCM/biblioteca/11979.php>. Accessed April 27, 2011.
- Bibliothèque Cantonale et Universitaire – Lausanne. n.d. "La BCU Et Google Livres ". <http://www.unil.ch/bcu/page74263.html>. Accessed January 17, 2012.

- Bijker, Wiebe. 1992. "The Social Construction of Fluorescent Lighting, or How an Artifact Was Invented in Its Diffusion Stage." In *Shaping Technology/Building Society: Studies in Sociotechnical Change*, edited by Wiebe Bijker and John Law. 75-102. Cambridge, MA: MIT Press.
- Bijker, Wiebe, Thomas P. Hughes, and Trevor Pinch. 1987. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Boston: MIT Press.
- Bijker, Wiebe, and John Law. 1992a. "General Introduction." In *Shaping Technology/Building Society: Studies in Sociotechnical Change*, edited by Wiebe Bijker and John Law. 1-14. Cambridge, MA: MIT Press.
- — —. 1992b. *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press.
- Biodiversity Heritage Library. n.d. "Scanning Processes."
<http://biodivlib.wikispaces.com/Scanning+Process>. Accessed January 22, 2012.
- Bishop, Ann Peterson, Nancy A. Van House, and Barbara P. Battenfield. 2003. *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA: MIT Press.
- Bivens-Tatum, Wayne. 2011. *Libraries and the Enlightenment*. Los Angeles: Library Juice Press.
- Black, Alistair, Simon Pepper, and Kaye Bagshaw. 2009. *Books, Buildings, and Social Engineering: Early Public Libraries in Britain from Past to Present*. Farnham: Ashgate.
- Board of Aldermen. 1857. "Communication from the Mayor, Transmitting a Memorial from the Trustees of the Public Library, Suggesting Certain Changes in the Library Ordinance." City Document No. 70. Boston. November 5. Courtesy Boston Public Library: Government Documents Department.
- Board of Trustees. 1853. *Proposed Rules and Regulations of the Public Library of the City of Boston*. Boston: J.H. Eastburn, City Printer. Courtesy Boston Public Library: Delivery Desk.
- — —. 1858. "Rules and Regulations of the Public Library of the City of Boston [Annotated by G. Ticknor]." August 17. Courtesy Boston Public Library: Delivery Desk.
- Bobinski, George S. 1969. *Carnegie Libraries*. Chicago: American Library Association.
- Boston City Council. 1858. "Proceedings at the Dedication of the Building for the Public Library of the City of Boston." Boston. January 1.
- Boston Common Council. 1852a. "An Ordinance in Relation to the Public Library." City Document No. 57. October 14. Courtesy Boston Public Library: Government Documents Department.
- — —. 1852b. "Report of Committee on Public Library." City Document No. 79. Courtesy Boston Public Library: Government Documents Department.
- — —. 1854a. "An Ordinance in Relation to the Public Library Commissioners." City Document No. 120. December 21. Courtesy Boston Public Library: Government Documents Department.
- — —. 1854b. "Report of Committee on Public Library." City Document No. 21. February 23. Courtesy Boston Public Library: Government Documents Department.
- "Boston Library Consortium and Open Content Alliance to Provide Digitized Books." 2007. *College & Research Libraries News* 68, no. 10: 624-25.
- "Boston Library Consortium, Open Content Alliance Partner." 2007. *Advanced Technology Libraries* 36, no. 11 (November): 1, 9-10.
- Boston Public Library. "Boston Public Library: A Brief History and Description."
<http://www.bpl.org/general/history.htm>. Accessed March 15, 2011.
- Boston School Committee. 1949. *Brief History of the Boston Public Schools, 1635-1949*. Boston.

- Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Inside Technology. Cambridge, MA: MIT Press.
- Courtesy N/A.
- "Brewster Kahle." 2013. NNDB. <http://www.nndb.com/people/831/000124459/>. Accessed August 27, 2013.
- British Library. 2011. "The British Library and Google to Make 250,000 Books Available to All." <http://pressandpolicy.bl.uk/Press-Releases/The-British-Library-and-Google-to-make-250-000-books-available-to-all-4fc.aspx>. Accessed July 8, 2013.
- Brundin, Robert E. 1975. "Justin Winsor of Harvard and the Liberalizing of the College Library." *The Journal of Library History* (1974-1987) 10, no. 1: 57-70.
- "Búsqueda De Libros Electrónicos UCM-Google." n.d. Biblioteca Complutense: Colección Digital Complutense. <http://www.ucm.es/BUCM/atencion/25403.php>. Accessed April 27, 2011.
- "A Call to Participate in the Open Content Alliance." n.d. <http://www.opencontentalliance.org/participate/>. Accessed February 2, 2012.
- Calter, Mimi. 2007. "Stanford Launches Copyright Renewal Database." *Stanford News Service*, April 2.
- Capen, Edward. 1900? "Sketch of the Early History of the Boston Public Library." Courtesy Boston Public Library: Rare Books & Manuscripts.
- Carey, James W. 1967. "Harold Adams Innis and Marshall McLuhan." *The Antioch Review* 27, no. 1: 5-39.
- Carlson, Scott, and Jeffrey R. Young. 2005. "Yahoo Works with Academic Libraries on a New Project to Digitize Books." *Chronicle of Higher Education* 52, no. 8 (October 14): A34.
- Carlson, W. Bernard. 1992. "Artifacts and Frames of Meaning: Thomas A. Edison, His Managers, and the Cultural Construction of Motion Pictures." In *Shaping Technology/Building Society: Studies in Sociotechnical Change*, edited by Wiebe Bijker and John Law. 175-98. Cambridge, MA: MIT Press.
- Carlyle, Allyson. 2010. "Content Analysis across the Disciplines: A Framework."
- Carnegie, Andrew. 1889a. "The Best Fields for Philanthropy." *The North American Review* 149, no. 397: 682-98.
- — —. 1889b. "Dedication of the Carnegie Library at the Edgar Thomson Steel Rail Works, Braddock: Address to the Workmen." March 30. Courtesy Library of Congress Manuscript Division.
- — —. 1889c. "Wealth." *The North American Review* 148, no. 391: 653-64.
- — —. 1890. "Presentation of the Freedom of the Burgh of Inverness to Andrew Carnegie, Esq." September 12. Courtesy Library of Congress Manuscript Division.
- — —. 1891. "Address by Mr. Carnegie." In *Souvenir: Laying the Foundation Stone of the Peterhead Free Library and Museum*. August 8. Courtesy Library of Congress Manuscript Division.
- — —. 1894. [Dedication of Jedburgh Library Building]. October. Courtesy Library of Congress Manuscript Division.
- — —. 1895. "Presentation of the Carnegie Library to the People of Pittsburgh, with a Description of the Dedicatory Exercises." November 5.
- — —. 1898a. [Founder's Day Address, Pittsburgh]. Courtesy Library of Congress Manuscript Division.
- — —. 1898b. "Homestead Address." Courtesy Library of Congress Manuscript Division.
- — —. 1909. "Brooklyn." [Speech]. Courtesy Library of Congress Manuscript Division.
- — —. 1920. *Autobiography of Andrew Carnegie*. London: Constable & Co. Limited.

- Carnegie Corporation of New York. n.d. "Free Public Library Bilding." Andrew Carnegie Collection. [Application form]. Courtesy Carnegie Mellon University Libraries, Digital Collections.
- Carnegie Endowment for International Peace. 2010. "Centennial of the Global Think Tank." <http://www.carnegieendowment.org/about/index.cfm?fa=centennial>. Accessed January 11, 2013.
- Carnegie Library of Pittsburgh. 1897. "First Annual Report to the Board of Trustees of the Carnegie Library of Pittsburgh." January 31.
- — —. 1898. "Second Annual Report to the Board of Trustees of the Carnegie Library of Pittsburgh." January 31.
- — —. 1899. "Third Annual Report to the Board of Trustees of the Carnegie Library of Pittsburgh." January 31.
- Carnegie Public Library, Ayr, and Andrew Carnegie. 1892. "Address by Andrew Carnegie, Esq., on the Occasion of Laying the Memorial Stone." October 5. Courtesy Library of Congress Manuscript Division.
- Carr, Reg. 2005a. "From Gutenberg to Google: The Case of the Bodleian Library, Oxford." Presentation to the Union League of Philadelphia. <http://www.bodley.ox.ac.uk/librarian/rpc/gg2gg/gg2gg.htm>. Accessed April 27, 2011.
- — —. 2005b. "The Oxford-Google Mass-Digitisation Programme." CNI Spring 2005 Task Force Meeting. <http://www.bodley.ox.ac.uk/librarian/rpc/CNIGoogle/CNIGoogle.htm>. Accessed April 27, 2011.
- — —. 2005c. "The Oxford-Google Mass-Digitisation Project: How, Why and What?". Educause Webcast. June 15.
- Ceynowa, Klaus. 2009. "Mass Digitization for Research and Study: The Digitization Strategy of the Bavarian State Library." *IFLA Journal* 35, no. 1: 17-24.
- Chen, Cindy. 2011. "Chronicling the Life of Chez Panisse Restaurant, One Menu at a Time." *Inside Google Books* (blog). October 27. <http://booksearch.blogspot.com/2011/07/first-google-ebooks-integrated-e-reader.html>. Accessed July 8, 2013.
- Chen, Xiaotian. 2012. "Google Books and Worldcat: A Comparison of Their Content." *Online Information Review* 36, no. 4 (2012): 507-16.
- Chillingworth, Mark. 2005. "Yahoo Challenges Google Book Digitisation Project." *Information World Review*, no. 217 (October).
- Chin, Denny. 2013. "Opinion." In *The Authors Guild, Inc., and Betty Miles, Joseph Goulden, and Jim Bouton, on behalf of themselves and all others similarly situated v. Google, Inc.*, 05 Civ. 8136 (DC).
- Choudhury, G. Sayeed, Tim DiLauro, Robert Ferguson, Michael Doretboom, and Ichiro Fujinaga. 2006. "Document Recognition for a Million Books." *D-Lib Magazine* 12, no. 3.
- Ciriaci, Daria, and Davide Quaglione. 2009. "Multilateral Platforms and Google Book Search: Which Competition after the Settlement Agreement?" [In Italian]. *Industria: Rivista di economia e politica industriale* 30, no. 4: 647-78.
- Clark, Alex. 2013. "Reclaiming New Zealand's Digitized Heritage: Key Findings to Date." *Kiwi Alex* (blog). March 31. <http://www.kiwialex.com/1/post/2013/03/about-the-project.html>. Accessed July 24, 2013.
- Cleland, Scott, and Ira Brodsky. 2011. *Search and Destroy: Why You Can't Trust Google Inc.* St. Louis, MO: Telescope Books.
- Cliatt, Cass. 2007. "Library Joins Google Project to Make Books Available Online." News at Princeton, Princeton University.

- <http://www.princeton.edu/main/news/archive/S16/84/71S02/index.xml>. Accessed April 27, 2011.
- Coleman, Mary Sue. 2006. "Google, the Khmer Rouge, and the Public Good." Address to the Professional/Scholarly Publishing Division of the Association of American Publishers. February 6.
- Colford, Michael R. 2008. "Rethinking Resource Sharing: Boston Public Library Provides Scan-on-Demand for Interlibrary Loan." *ASCLA* 30, no. 1: 5-6.
- Colombet, Philippe. 2010. "Google Books Goes Dutch." *Official Google Blog* (blog). July 14. <http://googleblog.blogspot.com/2010/07/google-books-goes-dutch.html>. Accessed July 8, 2013.
- Columbia University Libraries. 2007a. "Columbia University Libraries Becomes Newest Partner in Google Book Search Library Project." *Library News*. http://library.columbia.edu/news/libraries/2007/20071213_google.print.html. Accessed April 27, 2011.
- — —. 2007b. "Frequently Asked Questions About the Columbia University - Google Book Search Library Partnership." *Library News*. http://library.columbia.edu/news/libraries/2007/20071213_googlefaq.print.html. Accessed April 27, 2011.
- Colvin, Jennifer. 2006. "UC Libraries Partner with Google to Digitize Books." California Digital Library. <http://www.universityofcalifornia.edu/news/2006/aug09.html>. Accessed April 27, 2011.
- Committee on Institutional Cooperation. 2008. "CIC-Google FAQ." CIC - Google Book Search Project. <http://www.cic.net/Home/Projects/Library/BookSearch/FAQ.aspx>. Accessed April 27, 2011.
- — —. n.d. "Google Book Search Project - Introduction." CIC - Google Book Search Project. <http://www.cic.net/Home/Projects/Library/BookSearch/Introduction.aspx>. Accessed April 27, 2011.
- Committee on Institutional Cooperation, and Google, Inc. 2007. "Cooperative Agreement." CIC - Google Book Search Project. <http://www.cic.net/Home/Projects/Library/BookSearch/CIC-Google.aspx>. Accessed April 27, 2011.
- "Consortium Forms OCA to Bring Additional Content Online." 2005. *Advanced Technology Libraries* 34, no. 11: 9-10.
- Consortium of Academic and Research Libraries in Illinois. 2007. "CARLI Book Digitization Initiative."
- "Content Agreement." 2013. [Template]. San Francisco: Internet Archive. August 14.
- Cooper, Alan. 2004. *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity*. Indianapolis: Sams Publishing.
- Cornell University Library. n.d. "FAQ: Cornell University Library-Google Library Partnership." Library Communications. <http://www.library.cornell.edu/communications/Google/faq.html>. Accessed April 27, 2011.
- "Cornell University Library Is Newest Partner in Google Book Search Library Project." 2007. *Library Hi Tech News* 24, no. 9/10: 49.
- "Correspondence." 1890-1916. Andrew Carnegie Collection. [Search Term: "Carnegie Library"; Manual filtering to correspondence related to the library program]. Courtesy Carnegie Mellon University Libraries, Digital Collections.

- Courant, Paul N. 2006. "Scholarship and Academic Libraries (and Their Kin) in the World of Google." *First Monday* 11, no. 8.
- — —. 2008. "The Google Settlement - from the Universal Library to the Universal Bookstore." *Au Courant* (blog). October 28. <http://paulcourant.net/2008/10/28/the-google-settlement-from-the-universal-library-to-the-universal-bookstore/>. Accessed April 12, 2009.
- Courant, Paul N., and Matthew Nielsen. 2010. "On the Cost of Keeping a Book." In *The Idea of Order*. 81-105. Washington, DC: Council on Library and Information Resources.
- Coyle, Karen. 2006. "Mass Digitization of Books." *Journal of Academic Librarianship* 32, no. 6: 641-45.
- — —. 2008. "Amputation." *Coyle's InFormation* (blog). May 26. <http://kcoyle.blogspot.com/2008/05/amputation.html>. Accessed July 31, 2013.
- — —. 2009. "The Google/ AAP Settlement: What's in It for Libraries?" In *Panel Presentation at the ALA Midwinter Conference*. Denver, CO. Courtesy Repository.
- Crawford, James. 2010. "On the Future of Books." *Inside Google Books* (blog). October 14. <http://booksearch.blogspot.com/2010/10/on-future-of-books.html>. Accessed July 8, 2013.
- Crawford, Walt. 2005. "OCA and Glp 2: Steps on the Digitization Road." *Cites & Insights* 5, no. 14.
- — —. 2006. "Discovering Books: The OCA/GBS Saga Continues." *Cites & Insights* 6, no. 6 (Spring).
- — —. 2007. "Book Searching: OCA/GBS Update." *Cites & Insights* 7, no. 1.
- — —. 2009. "On Privatization." *Cites & Insights* 9, no. 8 (July).
- Cunningham, Kelly Eileen. 2008. "Major Universities See Promise in Google Book Search Settlement." October 28.
- Dana, John Cotton. 1897. "The Public and Its Public Library." *Appletons' Popular Science Monthly* 51: 242-53.
- Darnton, Robert. 2009. "Google & the Future of Books." *New York Review of Books* 56, no. 2 (February 12).
- — —. 2010a. "The Google Books Settlement: An Exchange." *New York Review of Books* 57, no. 3 (February): 40.
- — —. 2010b. "The Library: Three Jeremiads." *The New York Review of Books*, December 23.
- — —. 2011a. "A Digital Library Better Than Google's." *New York Times*, March 23.
- — —. 2011b. "Google's Loss: The Public's Gain." *New York Review of Books* 58, no. 7: 10-12.
- — —. 2013. "The National Digital Public Library Is Launched!". *New York Review of Books* (April 25).
- David, Paul A. 1985. "Clio and the Economics of Qwerty." *The American Economic Review* 75, no. 2 (May): 332-37.
- Davis, Janet M. 2002. *The Circus Age: Culture & Society under the American Big Top*. Chapel Hill: University of North Carolina Press.
- "Dedication of Washington (D.C.) Carnegie Library Building." 1903. *Library Journal* 28, no. 1 (January): 18.
- DeSanctis, Gerardine, and Marshall Scott Poole. 1994. "Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory." *Organization Science* 5, no. 2 (May): 121-47.
- Dewey, John. 1916. *Democracy and Education: An Introduction to the Philosophy of Education*. New York: Free Press. Reprint, 1966.
- Ditzion, Sidney. 1947. *Arsenals of a Democratic Culture*. Chicago: ALA.

- Documents Relating to a Meeting of Citizens of Boston Held May 5, 1841, to Consider Mr. A. Vattemare's Plan of a Public Literary and Scientific Institution [in Boston] and System of International Literary Exchange.* 1841. Boston. Courtesy Boston Public Library: Rare Books & Manuscripts.
- Dougan, Kirstin. 2010. "Music to Our Eyes: Google Books, Google Scholar, and the Open Content Alliance." *portal: Libraries & the Academy* 10, no. 1 (January): 75-93.
- Dougherty, William C. 2010. "The Google Books Project: Will It Make Libraries Obsolete?." *Journal of Academic Librarianship* 36, no. 1 (January): 86-89.
- Drummond, David. 2005. "Why We Believe in Google Print." *Google Blog* (blog). October 19. <http://googleblog.blogspot.com/2005/10/why-we-believe-in-google-print.html>.
- Dubé, Line, and Guy Paré. 2003. "Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations." *MIS Quarterly* 27, no. 4: 597-636.
- Duguid, Paul. 2007. "Inheritance and Loss? A Brief Survey of Google Books." *First Monday* 12, no. 8.
- Duncker, Elke. 2002. "Cross-Cultural Usability of the Library Metaphor." Paper presented at *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital libraries*, Portland, OR. ACM.
- Duriau, Vincent J., Rhonda K. Reger, and Michael D. Pfarrer. 2007. "A Content Analysis of the Content Analysis Literature in Organization Studies: Research Themes, Data Sources, and Methodological Refinements." *Organizational Research Methods* 10, no. 1 (January): 5-34.
- "Eighth Annual Report of the Trustees of the City Library." 1860. City Document No. 90. Boston: City of Boston. November 19.
- Erschine, Andrew. 1995. "Culture and Power in Ptolemaic Egypt: The Museum and Library of Alexandria." *Greece & Rome* 42, no. 1: 38-48.
- "The Experiment of Free Libraries." 1867. *The Round Table. A Saturday Review of Politics, Finance, Literature, Society and Art* 6, no. 151 (December 14): 392.
- Feng, Shaolei, and R. Manmatha. 2006. "A Hierarchical, HMM-Based Automatic Evaluation of OCR Accuracy for a Digital Library of Books." In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*. Chapel Hill, NC: ACM.
- "Fifth Annual Report of the Trustees of the City Library." 1857. City Document No. 71. Boston: City of Boston. November 10.
- "First Annual Report of the Carnegie Free Library, Allegheny, PA, for the Year Ending 28th of February, 1891." 1891. Oil City, PA: Derrick Publishing Company, Printers and Bookbinders. February 28.
- Foster, Andrea L. 2006. "Google Counters Critics of Library Project." *Chronicle of Higher Education* 52, no. 27: A32.
- "Fourth Annual Report of the Trustees of the City Library." 1856. City Document No. 61. Boston: City of Boston. October 31.
- Fox, Robert. 1996. "Technological Change: Methods and Themes in the History of Technology." *Studies in the History of Science, Technology and Medicine* 1.
- "Free Public Library: Atlanta, Georgia." 1898. Carnegie Corporation of New York Library Correspondence. Microfilm Reel 2. Courtesy Columbia University Rare Books and Manuscripts Library.
- "Free Public Library: Ballard, Washington." 1903. Carnegie Corporation of New York Library Correspondence. Microfilm Reel 2. Courtesy Columbia University Rare Books and Manuscripts Library.

- "Free Public Library: Carnegie, Pennsylvania." 1898. Carnegie Corporation of New York Library Correspondence. Microfilm Reel 5. Courtesy Columbia University Rare Books and Manuscripts Library.
- "Free Public Library: Johnstown, Pennsylvania." 1890. Carnegie Corporation of New York Library Correspondence. Microfilm Reel 15. Courtesy Columbia University Rare Books and Manuscripts Library.
- "Free Public Library: New York, New York." 1899. Carnegie Corporation of New York Library Correspondence. Microfilm Reel 22. Courtesy Columbia University Rare Books and Manuscripts Library.
- "Free Public Library: Olympia, Washington." 1903. Carnegie Corporation of New York Library Correspondence. Microfilm Reel 23. Courtesy Columbia University Rare Books and Manuscripts Library.
- "Free Public Library: Pittsburgh, Pennsylvania." 1890. Carnegie Corporation of New York Library Correspondence. Microfilm Reel 25. Courtesy Columbia University Rare Books and Manuscripts Library.
- "Frequently Asked Questions About the Million Book Project." 2007.
https://libwebspace.library.cmu.edu/libraries-and-collections/MBP_FAQ.html#current. Accessed August 21, 2013.
- Frueh, Linda. 2008. "New Contributors to Open Content Alliance Collections." *Opencontentalliance.org* (blog). November 9.
<http://www.opencontentalliance.org/2008/11/09/new-contributors-to-open-content-alliance-collections/#more-241>. Accessed February 2, 2012.
- Ganz, Cheryl. 2008. *The 1933 Chicago World's Fair: Century of Progress*. Urbana: University of Illinois Press.
- Garrison, Dee. 1973. "The Tender Technicians: The Feminization of Public Librarianship, 1876-1905." *Journal of Social History* 6, no. 2: 131-59.
- — —. 1976. "Immoral Fiction in the Late Victorian Library." *American Quarterly* 28: 71-89.
- — —. 1979. *Apostles of Culture: The Public Librarian and American Society, 1876-1920*. New York: Macmillan.
- Georg, Johann, of Saxony, and E. Daenell, eds. 1920. *Briefwechsel König Johanns Von Sachsen Mit George Ticknor*. Leipzig and Berlin: B. G. Teubner.
- Ghent University Library. 2007. "Google and Ghent University Library to Make Hundreds of Thousands of Dutch and French Books Available Online."
<http://lib.ugent.be/files/en/pdf/20070523-persnota-en.pdf>. Accessed April 27, 2011.
- — —. 2009. "Ghent - Google - Treasures." August 28.
- Giddens, Anthony. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California Press.
- Giersberg, Dagmar. 2007. "The Bavarian State Library' Books Are Being Digitalised by Google." Goethe-Institut. <http://www.goethe.de/wis/bib/dib/en2332286.htm>. Accessed April 27, 2011.
- Gleason, Eliza Atkins. 1941. "Public Library Service and the Negro." Paper presented at *Library Conference*, Atlanta, GA, March 14-15. Courtesy Columbia University Rare Books and Manuscripts Library.
- Glushko, Bobby. 2012. "CRMS-World: Legal Analysis and Resulting Workflow." Ann Arbor: University of Michigan Copyright Office. February 20.
- Gohring, Nancy. 2005. "Google Print Gets New Name." *InfoWorld* (17 November).
- Goldsborough, Reid. 2006. "Searching the Full Text of Books." *Information Today* 23, no. 1: 29-30.

- — —. 2008. "Led by Google, Internet Access to Books Making Serious Strides." *Community College Week* 20, no. 10: 17.
- Gonsalves, Antone. 2006. "Internet Archive Claims Progress against Google Library Initiative." *InformationWeek*, December 20.
- Google, Inc. 2004. "Google Checks out Library Books." http://www.google.com/press/pressrel/print_library.html. Accessed June 8, 2008.
- — —. 2007. "Google Books History." <http://www.google.com/googlebooks/about/history.html>. Accessed June 11, 2013.
- — —. 2011a. "Google Books Library Project - an Enhanced Card Catalog of the World's Books." <http://books.google.com/googlebooks/library.html>. Accessed March 7, 2011.
- — —. 2011b. "Google Books Terms of Service." <http://books.google.com/intl/en/googlebooks/tos.html>. Accessed May 1, 2012.
- — —. 2011c. "What You'll See When You Search on Google Books." <http://books.google.com/intl/en/googlebooks/screenshots.html>. Accessed November 11, 2011.
- — —. 2012a. "How to Claim Books That Were Scanned for the Library Project." Google Books Help. https://support.google.com/books/partner/answer/2520009?hl=en&ref_topic=2376976. Accessed July 23, 2013.
- — —. 2012b. "Remove a Book from Google Books." Google Books Help. <https://support.google.com/books/answer/43756?hl=en>. Accessed July 23, 2013.
- — —. n.d.-a. "About Search." Facts about Google and Competition. <http://www.google.com/competition/howgooglesearchworks.html>. Accessed June 27, 2013.
- — —. n.d.-b. "Official Blog." <http://googleblog.blogspot.com>. Accessed June 17, 2013.
- — —. n.d.-c. "Partner Program Tour." <http://www.google.com/googlebooks/partners/tour.html>. Accessed June 27, 2013.
- "Google Print." 2003. *Kottke.org* (blog). December 16. <http://kottke.org/03/12/google-print>. Accessed June 7, 2013.
- "Google Print Goes Live." 2004. *Publishers Weekly* (6 October).
- Goth, G. 2007. "Digital Libraries Are Taking Form." *IEEE Distributed Systems Online* 8, no. 12: 1-3.
- Graf, Klaus. 2006. "Google Books Is Blocking Users Outside the US." *Archivalia* (blog). November 12. <http://archiv.twoday.net/stories/2922570/>. Accessed July 23, 2013.
- Graham, Jefferson. 2005. "Google Print Project Inspires Fans, Fears." *USA Today*.
- Granka, Laura A. 2010. "The Politics of Search: A Decade Retrospective." *The Information Society* 26, no. 5 (2010/09/27): 364-74.
- Green, Samuel Swett. 1876. "Personal Relations between Librarians and Readers." *Library Journal* 1 (October).
- Greenough, William W. 1891? "Annals of the Public Library of the City of Boston from the Year 1848 to Its Removal to the New Edifice in Copley Square." Boston, MA. Courtesy Boston Public Library: Rare Books & Manuscripts.
- Greenstein, Daniel, and Suzanne E. Thorin. 2002. "The Digital Library: A Biography."
- Grimmelman, James. 2010. "GBS: Internet Archive Starts Lending in-Copyright E-Books." *The Laboratorium* (blog). June 30. http://laboratorium.net/archive/2010/06/30/gbs_internet_archive_starts_lending_in-copyright_e. Accessed September 19, 2013.

- Grimmelmann, James. 2009a. "The Google Book Search Settlement: Ends, Means, and the Future of Books."
- — —. 2009b. "Why It Matters." *Publishers Weekly*, 20-21.
- Grogg, Jill E., and Beth Ashmore. 2007. "Google Book Search Libraries and Their Digital Copies: What Now?". *Searcher* 15, no. 4 (April): 18-21, 23-7.
- Guess, Andy. 2008. "Post-Microsoft, Libraries Mull Digitization." *Inside Higher Ed*, May 30.
- Guren, Cliff. 2006. "Live Search Books Beta Release." *Bing Community* (blog). December 5. http://www.bing.com/community/site_blogs/b/search/archive/2006/12/05/live-search-books-beta-release.aspx.
- — —. 2007. "Live Search Books: Now with in-Copyright Content." *Bing Community* (blog). May 31. http://www.bing.com/community/site_blogs/b/search/archive/2007/06/01/live-search-books-now-with-in-copyright-content.aspx. Accessed February 2, 2012.
- Hadro, Josh. 2009. "Kirtas and Penn Team up to Scan." *Library Journal* 134, no. 5 (March 15): 20.
- — —. 2010a. "Infotech." *Library Journal* 135, no. 5 (March 15): 16.
- — —. 2010b. "Infotech." *Library Journal* 135, no. 10 (June 1): 18.
- Hafner, Katie. 2005. "In Challenge to Google, Yahoo Will Scan Books." *New York Times*, October 3, C1(L).
- Hahn, Trudi Bellardo. 2008. "Mass Digitization: Implications for Preserving the Scholarly Record." *Library Resources & Technical Services* 52, no. 1 (January): 18-26.
- Hane, Paula J. 2005. "Digitization Plans Top the News." *Information Today* 22, no. 10: 7-14.
- — —. 2007. "Free Content Options Continue to Shake Things Up." *Information Today* 24, no. 11: 7-12.
- Hardy, Quentin. 2009. "The Big Deal: Brewster Kahle." *Forbes*, November 27.
- Harris, Michael H. 1975a. "The Role of the Public Library in American Life: A Speculative Essay." *Occasional Papers* (January).
- — —. 1975b. "The Role of the Public Library in American Life: A Speculative Essay." *Occasional Papers*, 117. Champaign, IL: University of Illinois Graduate School of Library Science. January.
- — —. 1976. "Public Libraries and the Decline of the Democratic Dogma." *Library Journal* 101, no. 19: 2225-30.
- — —. 1986. "State, Class, and Cultural Reproduction: Toward a Theory of Library Service in the United States." *Advances in Librarianship* 14: 211-52.
- — —. 1995. *History of Libraries in the Western World*. Metuchen, NJ: Scarecrow Press.
- Harrison, Helen A., Joseph P. Cusker, and Queens Museum. 1980. *Dawn of a New Day: The New York World's Fair, 1939/40*. Flushing, NY: Queens Museum.
- Harvard University Library. 2004. "Harvard's Digitization Project with Google: Harvard Libraries and Google Announce Pilot Digitization Project with Potential Benefits to Scholars Worldwide." Harvard University Library News Archive. http://hul.harvard.edu/news/2004_1214_news.html. Accessed April 27, 2011.
- HathiTrust Digital Library. 2012. "Ten Million and Counting." <http://www.hathitrust.org/blogs/perspectives-from-hathitrust/ten-million-and-counting>. Accessed July 4, 2013.
- — —. 2013. "HathiTrust Languages." http://www.hathitrust.org/visualizations_languages. Accessed July 14, 2013.
- — —. n.d.-a. "Help - Copyright." www.hathitrust.org/help_copyright. Accessed June 28, 2013.
- — —. n.d.-b. "Partnership Community." <http://www.hathitrust.org/community>. Accessed July 24, 2013.

- "Have a Hand in Scan-on-Demand." 2008. *Library Journal* 133, no. 20 (December 15): 27.
- Heald, Paul J. 2013. "How Copyright Makes Books and Music Disappear (and How Secondary Liability Rules Help Resurrect Old Songs)." Illinois Program in Law, Behavior and Social Science Paper. July 5.
- Hein, Virginia H. 1972. "The Image of 'a City Too Busy to Hate': Atlanta in the 1960's." *Phylon* 33, no. 3: 205-21.
- Hellman, Eric. 2010. "Internet Archive Sets Fair-Use Bait with Open Library Lending." *Go to Hellman* (blog). July 2. <http://go-to-hellman.blogspot.com/2010/07/internet-archive-sets-fair-use-bait.html>. Accessed September 19, 2013.
- Helm, Burt. 2005a. "Google's Escalating Book Battle." *BusinessWeek*: N.PAG.
- — —. 2005b. "Google's Great Works in Progress." *BusinessWeek*, December 22, 5.
- Hendrick, Burton. September 24-28, 1927. Manuscript Division, Library of Congress.
- — —. 1932. *The Life of Andrew Carnegie*. Vol. 1. Garden City, NY: Doubleday, Doran & Company.
- Heracleous, Loizos, and Michael Barrett. 2001. "Organizational Change as Discourse: Communicative Actions and Deep Structures in the Context of Information Technology Implementation." *The Academy of Management Journal* 44, no. 4: 755-78.
- Hill, Frank Pierce. 1936. *James Bertram: An Appreciation*. Carnegie Corporation of New York.
- Hillard, George S., Anna Ticknor, and Anna E. Ticknor, eds. 1876. *Life, Letters, and Journals of George Ticknor*. 5th ed. Vol. 2. Boston: J.R. Osgood and Company.
- Hobhouse, Christopher. 1937. *1851 and the Crystal Palace; Being an Account of the Great Exhibition and Its Contents; of Sir Joseph Paxton; and of the Erection, the Subsequent History and the Destruction of His Masterpiece*. New York: E.P. Dutton & Company, Inc.
- Hogge, Becky. 2005. "Brewster Kahle." *New Statesman* 134, no. 4762: 26.
- Holahan, Catherine. 2007. "Microsoft's Copyright Assault on Google." *BusinessWeek* (7 March).
- Holmes, David. 2005. *Communication Theory: Media, Technology, and Society*. London: Sage.
- Howard, Jennifer. 2012. "Google Begins to Scale Back Its Scanning of Books from University Libraries." *Chronicle of Higher Education* (March 9).
- Hughes, Thomas. 1987. "The Evolution of Large Technological Systems." In *The Social Construction of Technological Systems*, edited by Wiebe Bijker, Thomas Hughes and Trevor Pinch. 51-82. Cambridge, MA: MIT Press.
- Innis, Harold Adams. 1951. *The Bias of Communication*. Toronto: University of Toronto Press.
- International Telecommunications Union. 2013. "The World in 2013: ICT Facts and Figures." Geneva: International Telecommunications Union.
- Internet Archive. 2001. "Internet Archive's Terms of Use, Privacy Policy, and Copyright Policy." <http://archive.org/about/terms.php>. Accessed May 1, 2012.
- — —. 2007. "Internet Archive Scribe2." *Launchpad.net*. Accessed August 21, 2013.
- — —. n.d.-a. "About the Internet Archive." <https://archive.org/about/>. Accessed December 14, 2013.
- — —. n.d.-b. "Ebook and Texts Archive." <http://www.archive.org/details/texts>. Accessed September 17, 2013.
- — —. n.d.-c. "Frequently Asked Questions: Rights." <http://archive.org/about/faqs.php#Rights>. Accessed September 30, 2013.
- — —. n.d.-d. "Frequently Asked Questions: Texts and Books." http://archive.org/about/faqs.php#Texts_and_Books. Accessed May 1, 2012.
- Internet Archive, and Boston Public Library. 2007. "The John Adams Library Collection - Cooperative Agreement." April 13.

- Internet Archive, and University of California. 2006. "IA/UC Library Book Scanning Cooperation Agreement." August 23.
- "The Internet Archive Receives Grant from Alfred P. Sloan Foundation to Digitize and Provide Open Online Access to Historical Collections from Five Major Libraries." 2006. December 20.
- Jackson, Millie. 2008. "Using Metadata to Discover the Buried Treasure in Google Book Search." *Journal of Library Administration* 47, no. 1/2: 165-73.
- Jacsó, Péter. 2008. "Amazon, Google Book Search, and Google Scholar." *Online* 32, no. 2: 51-54.
- Janes, Joseph. 2003. *Introduction to Reference Work in the Digital Age*. New York: Neal-Schuman Publishers.
- Jeanneney, Jean-Noël. 2006. *Google and the Myth of Universal Knowledge: A View from Europe*. Chicago: University of Chicago Press.
- Johnson, Alvin S. 1916. "A Report to Carnegie Corporation of New York on the Policy of Donations to Free Public Libraries." New York: Carnegie Corporation of New York. November 18. Courtesy Columbia University Rare Books and Manuscripts Library.
- Johnson, Elmer D. 1970. *History of Libraries in the Western World*. 2nd ed. Metuchen, NJ: Scarecrow Press.
- Johnson, Richard K. 2007. "In Google's Broad Wake: Taking Responsibility for Shaping the Global Digital Library." *ARL Bimonthly Report* 250: 1-15.
- Jones, Edgar. 2010. "Google Books as a General Research Collection." *Library Resources & Technical Services* 54, no. 2 (April): 77-89.
- Jones, Elisabeth. 2006. "Pushing the Digitization Dialogue Forward: "Scholarship and Libraries in Transition" at the University of Michigan." *JEP: the Journal of Electronic Publishing* 9, no. 2 (June).
- Jones, Elisabeth A. 2010. "Conceptualizing Large-Scale Information Access Efforts: The Case for Historical Context." Paper presented at *ASIS&T 2010*, Pittsburgh, PA, October 22-27.
- — —. 2013a. Chapter 3 (eponymous). In *Library 2020: Today's Leading Visionaries Describe Tomorrow's Library*, edited by Joseph W. Janes. 15-23. Lanham, MD: Rowman and Littlefield.
- — —. 2013b. "'Sometimes I've Believed as Many as Six Impossible Things before Breakfast:' Outsider Naiveté as an Asset in the Initiation of Massive-Scale Book Provision Efforts." Paper presented at *iConference 2013*, Fort Worth, TX, February 12-15.
- — —. 2013c. "What We Talk About When We Talk About the Google Books Fair Use Decision." *Elisabeth Jones: Pondering Libraries, Technology, Media...* (blog). November 20. <http://elisabethjones.wordpress.com/2013/11/20/what-we-talk-about-when-we-talk-about-the-google-books-fair-use-decision/>. Accessed March 21, 2014.
- Jones, Elisabeth A., and Joseph W. Janes. 2010. "Anonymity in a World of Digital Books: Google Books, Privacy, and the Freedom to Read." *Policy & Internet* 2, no. 4.
- Jones, Elisabeth A., and Joseph T. Tennis. 2012. "Facets of Access: A Typology of Information Dissemination Systems." *Proceedings of the 2012 iConference*. Toronto, ON. ACM.
- Jones, Philip. 2009. "Google: UK Publishers Named as Plaintiffs." *Bookseller* (16 November).
- Jones, Theodore. 1997. *Carnegie Libraries across America: A Public Legacy*. New York: John Wiley & Sons, Inc.
- JTW. 2009. "Digital Library Controversy: Google Gives Ground at EU Hearing." *Der Spiegel* (11 September).
- Jung, Kihyun. 2011. "Google Ebooks Opens a New Chapter for Canadian Readers." *Inside Google Books* (blog). November 1. <http://booksearch.blogspot.com/2011/11/google-ebooks-opens-new-chapter-for.html>. Accessed July 8, 2013.

- Kahle, Brewster. 2005a. "Announcing the Open Content Alliance." *Yahoo! Search Blog* (blog). October 2. <http://www.ysearchblog.com/2005/10/02/announcing-the-open-content-alliance/>.
- — —. 2005b. "Bookscanning Launch and Vision of an Open Library." <http://www.archive.org/post/45685/bookscanning-launch-and-vision-of-an-open-library>. Accessed February 2, 2012.
- — —. 2005c. "Open Content Alliance: Call for Participation." <http://www.archive.org/post/44311/open-content-alliance-call-for-participation>. Accessed February 2, 2012.
- — —. 2007a. "Internet Archive Officially a Library." *Internet Archive Blogs* (blog). June 25. <http://blog.archive.org/2007/06/25/internet-archive-officially-a-library>. Accessed August 21, 2013.
- — —. 2007b. *Libraries Going Open*. San Francisco: Internet Archive.
- — —. 2008. "Books Scanning to Be Publicly Funded." <http://www.archive.org/post/194217/books-scanning-to-be-publicly-funded>. Accessed February 2, 2012.
- — —. 2009a. "60,000 Books from Library of Congress Go Online." <http://www.archive.org/post/283352/60000-books-from-library-of-congress-go-online>. Accessed February 2, 2012.
- — —. 2009b. "Achievements for Humanity." *Opencontentalliance.org* (blog). July 5. <http://www.opencontentalliance.org/2009/07/05/achievements-for-humanity/>. Accessed February 2, 2012.
- — —. 2009c. "Economics of Book Digitization." *Opencontentalliance.org* (blog). March 22. <http://www.opencontentalliance.org/2009/03/22/economics-of-book-digitization/>. Accessed February 2, 2012.
- — —. 2010a. "70,000 Books from Cornell Libraries Online Now." <http://www.archive.org/post/283343/70000-books-from-cornell-libraries-online-now>. Accessed February 2, 2012.
- — —. 2010b. "Digital Lending Library." <http://www.archive.org/post/312815/digital-lending-library>. Accessed February 2, 2012.
- — —. 2010c. "Milestone Reached! 250,000 Books from Utoronto." <http://www.archive.org/post/326578/milestone-reached-250000-books-from-utoronto>. Accessed February 2, 2012.
- — —. 2010d. "Over 1 Million Digital Books Now Available Free to the Print-Disabled." *Opencontentalliance.org* (blog). May 6. <http://www.opencontentalliance.org/2010/05/06/over-1-million-digital-books-now-available-free-to-the-print-disabled/>. Accessed August 21, 2013.
- — —. Interviewed by Elisabeth A. Jones, September 7, 2011. In person, at the Internet Archive, San Francisco, CA.
- Karle-Zenith, Anne. 2006. "Google Book Search and the University of Michigan." Paper presented at *26th Annual Charleston Conference*, Charleston, November 8-11.
- Kaufman, Peter B., and Jeff Ubois. 2007. "Good Terms - Improving Commercial-Noncommercial Partnerships for Mass Digitization: A Report Prepared by Intelligent Television for RLG Programs, OCLC Programs and Research." *D-Lib Magazine* 13, no. 11-12.
- Keio University. 2007. "Keio University to Partner with Google, Inc. For Digitalization and Release of Its Library Collection to the World for 'Formation of Knowledge of the Digital Era'." http://www.keio.ac.jp/english/press_release/070706e.pdf. Accessed April 27, 2011.

- Keller, Michael. 2011. "Stanford Initial Statement on Decision on Amended Settlement Agreement of the Google Library Project Suit." SU LAIR, Stanford University. <https://lib.stanford.edu/google-books/stanford-initial-statement-decision-amended-settlement-agreement-google-library-project>. Accessed April 27, 2011.
- Kelley, Michael. 2011. "Newsdesk." *Library Journal* 136, no. 20 (December 1): 14.
- Kilker, Julian, and Geri Gay. 1998. "The Social Construction of a Digital Library: A Case Study Examining Implications for Evaluation." *Information Technology and Libraries* 17, no. 2: 60-70.
- Kirby, Charles K. 1857. *Plans of Public Library*. Courtesy Boston Public Library Fine Arts Department.
- Klein, Hans K., and Daniel Lee Kleinman. 2002. "The Social Construction of Technology: Structural Considerations." *Science, Technology, & Human Values* 27, no. 1 (Winter): 28-52.
- Kniffel, Leonard. 2008. "Backed by Internet Archive, Entrepreneur Takes on OCLC." *American Libraries* (April).
- Koch, Theodore Wesley. 1917. *A Book of Carnegie Libraries*. White Plains, NY: The H.W. Wilson Company.
- Kousha, Kayvan, and Mike Thelwall. 2009. "Google Book Search: Citation Analysis for Social Science and the Humanities." *Journal of the American Society for Information Science & Technology* 60, no. 8: 1537-49.
- Kramer, Heather Isbell. 2012. "Eliminating the Hobby from Genealogy." *The Geneabrarian Reference Desk* (blog). <http://geneabrarian.blogspot.com/2012/01/eliminating-hobby-from-genealogy.html>. Accessed January 6,
- Krathwohl, David R. 1998. *Methods of Educational and Social Science Research: An Integrated Approach*. 2nd ed. Long Grove, IL: Waveland Press.
- Kupferschmid, Keith. 2005. "Are Authors and Publishers Getting Scroogled? (Cover Story)." *Information Today* 22, no. 11: 1-5.
- "La BCU De Lausanne Partenaire De Google Book Search." 2007. *bibliothecaire?* (blog). May 18. <https://bibliotheaire.wordpress.com/2007/05/18/la-bcu-de-lausanne-partenaire-de-google-book-search/>. Accessed July 8, 2013.
- "La Literatura Catalana S'obre Al Món Digital." 2007. La Xarxa. <http://www.laxarxa.com/noticia/la-literatura-catalana-s-obre-al-mon-digital>. Accessed July 8, 2013.
- Lackie, Robert J. 2005. "Google's Print and Scholar Initiatives: The Value of and Impact on Libraries and Information Services." *Internet Reference Services Quarterly* 10, no. 3/4: 57-70.
- — —. 2008. "From Google Print to Google Book Search: The Controversial Initiative and Its Impact on Other Remarkable Digitization Projects." *Reference Librarian* 49, no. 1: 35-53.
- LaFollette, Marcel C., and Jeffrey K. Stine. 1991. *Technology and Choice: Readings from Technology and Culture*. Chicago: University of Chicago Press.
- Langley, Adam, and Dan S. Bloomberg. 2006. "Google Books: Making the Public Domain Universally Accessible." Paper presented at *Document Recognition and Retrieval XIV*, San Jose, CA.
- — —. 2008. "Google Books: Making the Public Domain Universally Accessible." *Proceedings of the SPIE - The International Society for Optical Engineering*.
- Last, Jonathan V. 2007. "Google and Its Enemies." *Weekly Standard*, December 10, 16-21.
- Latour, Bruno. 1991. "Technology Is Society Made Durable." In *A Sociology of Monsters: Essays on Power, Technology, and Domination*, edited by John Law. 103-31. London: Routledge.

- — —. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- Lavoie, Brian, Lynn Silipigni Connaway, and Lorcan Dempsey. 2005. "Anatomy of Aggregate Collections: The Example of Google Print for Libraries." *D-Lib Magazine* 11, no. 9.
- Lavoie, Brian, and Roger C. Schonfeld. 2006. "Books without Boundaries: A Brief Tour of the System-Wide Print Book Collection." *Journal of Electronic Publishing* 9, no. 2 (Summer).
- Learned, William S. 1924. *The American Public Library and the Diffusion of Knowledge*. New York: Harcourt, Brace and Company.
- Lee, Robert. 1966. *Continuing Education for Adults through the American Public Library, 1833-1964*. Chicago: American Library Association.
- — —. 1971. "The People's University--the Educational Objective of the Public Library." *Reader in American Library History*: 117-24.
- Leetaru, Kalev. 2008. "Mass Book Digitization: The Deeper Story of Google Books and the Open Content Alliance." *First Monday* 13, no. 10 (October 6).
- Lefevre, Francois-Marie, and Marin Saric. 2009. Detection of Grooves in Scanned Images. US Patent 7,508,978 B1, issued March 24, 2009.
- "Sovereign Immunity." 2010. *Wex*. Ithaca, NY: Cornell University Law School.
- Lenssen, Philipp. 2007. "Freeing Google Books." *Google Blogoscoped* (blog). January 10. <http://blogoscoped.com/archive/2007-01-10-n14.html>. Accessed July 24, 2013.
- Lerner, Fred. 1998. *The Story of Libraries: From the Invention of Writing to the Computer Age*. New York: Continuum.
- Lessig, Lawrence. 2006. *Code Version 2.0*. New York: Basic Books.
- — —. 2010. "For the Love of Culture." *New Republic*, January 26.
- Levy, David M. 1994. "Fixed or Fluid? Document Stability and New Media." *European Conference on Hypertext* (September): 24-31.
- — —. 2000. "Digital Libraries and the Problem of Purpose." *Bulletin of the American Society for Information Science* 26, no. 6 (August/September): 22-25.
- — —. 2001. *Scrolling Forward: Making Sense of Documents in the Digital Age*. New York: Arcade.
- — —. 2003. "Documents and Libraries: A Sociotechnical Perspective." In *Digital Library Use: Social Practice in Design and Evaluation*, edited by Ann Peterson Bishop, Nancy A. Van House and Barbara P. Battenfield. 25-42. Cambridge, MA: MIT Press.
- Levy, Steven. 2011. *In the Plex: How Google Thinks, Works, and Shapes Our Lives*. New York: Simon & Schuster.
- Library of Congress. 2008. "Federal Scanning Center at the Library of Congress (Fedscan): Frequently Asked Questions." October 29.
- — —. n.d. "About the Collections." American Memory, Library of Congress. <http://memory.loc.gov/ammem/about/about.html>. Accessed December 15, 2013.
- Litman, Jessica. 2001. *Digital Copyright: Protecting Intellectual Property on the Internet*. Amherst, N.Y.: Prometheus Books.
- Lorenzen, Michael. 1999. "Deconstructing the Carnegie Libraries: The Sociological Reasons Behind Carnegie's Millions to Public Libraries." *Illinois Libraries* 81, no. 2 (Spring): 75-78.
- Lu, Xiaonan, Brewster Kahle, James Z. Wang, and C. Lee Giles. 2008. "A Metadata Generation System for Scanned Scientific Volumes." Paper presented at *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, Pittsburgh, PA. ACM.
- Lydenberg, Harry Miller. 1916. "A History of the New York Public Library." *Bulletin of the New York Public Library, Astor, Lenox and Tilden Foundations* 20, no. 7 (July): 555-84.

- Lynch, Clifford A. 2006. "Closing Remarks." Paper presented at *Scholarship and Libraries in Transition: A Dialogue about the Impacts of Mass Digitization Projects*, Ann Arbor, MI, March 11.
- Lyons, Mary E. 2011. "It Isn't Ok to Steal." *Library Journal* 136, no. 20 (December 1): 10.
- Macleod, David I. 1968. *Carnegie Libraries in Wisconsin*. Madison: State Historical Society of Wisconsin for Dept. of History, University of Wisconsin.
- Madrigal, Alexis. 2010. "Inside the Google Books Algorithm." *The Atlantic*, November 1.
- Magdy, Walid, and Kareem Darwish. 2008. "Book Search: Indexing the Valuable Parts." Paper presented at *Proceeding of the 2008 ACM workshop on Research advances in large digital book repositories*, Napa Valley, CA. ACM.
- Mahoney, Honor. 2009. "Europe's Heated Reaction to Google Books." *BusinessWeek* (9 September).
- Maines, David. 2000. "The Social Construction of Meaning." *Contemporary Sociology* 29, no. 4 (July): 577-84.
- Malone, Cheryl Knott. 2000a. "Books for Black Children: Public Library Collections in Louisville and Nashville, 1915-1925." *The Library Quarterly* 70, no. 2: 179-200.
- — —. 2000b. "Toward a Multicultural American Public Library History." *Libraries & Culture* 35, no. 1 (Winter): 77-87.
- Mann, Thomas. 2008. "Will Google's Keyword Searching Eliminate the Need for LC Cataloging and Classification?". *Journal of Library Metadata* 8, no. 2: 159-68.
- Markoff, John, and Edward Wyatt. 2004. "Google Is Adding Major Libraries to Its Database." *New York Times*, 14 December, A1.
- Marsh, Ellen. 2007. "Cornell University Library Becomes Newest Partner in Google Book Search Library Project." Library Communications.
<http://www.library.cornell.edu/communications/Google/>. Accessed April 27, 2011.
- Martin, Robert S. 1993. *Carnegie Denied: Communities Rejecting Carnegie Library Construction Grants, 1898-1925*. Westport, CT: Greenwood Press.
- — —. 2007. "Intersecting Missions, Converging Practice." *RBM* 8, no. 1 (Spring): 80-88.
- Martin, Shawn. 2008. "To Google or Not to Google, That Is the Question: Supplementing Google Book Search to Make It More Useful for Scholarship." *Journal of Library Administration* 47, no. 1/2: 141-50.
- Mattiuzzo, Gino. 2010. "A Digital Renaissance: Partnering with the Italian Ministry of Cultural Heritage." *Official Google Blog* (blog). March 10.
<http://googleblog.blogspot.com/2010/03/digital-renaissance-partnering-with.html>. Accessed July 8, 2013.
- Maxymuk, John. 2005. "13 Ways of Looking at Googlelibrary." *Bottom Line: Managing Library Finances* 18, no. 4: 197-201.
- McCoy, Adrian. 2007. "The Internet Gives Birth to an 'Official' Online Library." *Pittsburgh Post-Gazette*, June 22.
- McCoy, Bill. 2007. "Google: A Glass-House Dweller Re: Misleading Notices?" *Bill McCoy: Just another Adobe Blogs site* (blog). August 1.
http://blogs.adobe.com/billmccoy/2007/08/google_a_glassh.html. Accessed July 24, 2013.
- McShane, Clay. 2008. "'Mind of God'-Not: Google Books for Urban Historians." *Journal of Urban History* 35, no. 1: 178-85.
- Meadows, Chris. 2013. "The Internet Archive's Open Library Is Violating Authors' Copyrights." *TeleRead* (blog). July 10. <http://www.teleread.com/copy-right/the-internet-archives-open-library-is-violating-authors-copyrights/>. Accessed September 19, 2013.

- "Members." 2009. Open Book Alliance. <http://www.openbookalliance.org/members/>. Accessed September 5, 2013.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, *et al.* 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331, no. 6014 (January 14, 2011): 176-82.
- "Michigan Digitization Project (Google) Timeline." n.d. Internal Planning Document. University of Michigan. Courtesy University of Michigan Libraries.
- "Microsoft Partners with Yahoo to Digitise Libraries." 2005. *Information World Review*, no. 218: 3-3.
- "Microsoft Scans British Library." 2005. *BBC News*, November 4.
- Miles, Matthew B., and A. Michael Huberman. 1984. *Qualitative Data Analysis: A Sourcebook of New Methods*. Beverly Hills, CA: Sage.
- "Milestone Achieved." 2006. *Opencontentalliance.org* (blog). December 20. <http://www.opencontentalliance.org/2006/12/20/milestone-achieved/>. Accessed February 2, 2012.
- Miller, Durand R. 1943. *Carnegie Grants for Library Buildings, 1890-1917. A List of Library Buildings, Public and Academic, Erected with Funds Provided by Andrew Carnegie and Carnegie Corporation of New York*. New York: Carnegie Corporation of New York.
- Miller, Robert. 2010. "Saveusjobspasssenatebills4213." Internet Archive. <http://archive.org/details/SaveUSJobsPassSenateBills4213>. Accessed August 21, 2013.
- — —. 2012. "Internet Archive Book Digitization Process." Rev. W. San Francisco: Internet Archive. July 10.
- Milliot, Jim. 2007. "Google's Book Search Moves Forward." *Publishers Weekly*, May 25, 14.
- Milliot, Jim, and Andrew Albanese. 2004. "Google Deal with Libraries Wins Praise, Raises Questions." *Publishers Weekly*, 7.
- Milliot, Jim, and Steve Zeitchik. 2004. "Google Begins to Push Google Print Program." *Publishers Weekly*, October 8.
- Milne, Ronald. 2005. "The Google Library Project at Oxford." *Internet Reference Services Quarterly* 10, no. 3/4: 23-28.
- — —. 2006. "The Google Mass Digitisation Project at Oxford." *Liber Quarterly: The Journal of European Research Libraries* 16, no. 1-4: 265-70.
- Mimno, David, and Andrew McCallum. 2007. "Organizing the OCA: Learning Faceted Subjects from a Library of Digital Books." Paper presented at *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, Vancouver, BC. ACM.
- MLibrary. 2008. "One Million Digitized Books." <http://www.lib.umich.edu/one-million-digitized-books>. Accessed June 14, 2013.
- "Mr. Andrew Carnegie in Aberdeen. Opening of the Public Library. Presentation of the Freedom of the City." 1892. In *Souvenir: Laying the Foundation Stone of the Peterhead Free Library and Museum*. July 6. Courtesy Library of Congress Manuscript Division.
- MSN. 2005. "MSN Search Announces MSN Book Search." October 25.
- "MSN Search Announces MSN Book Search." 2005. *Advanced Technology Libraries* 34, no. 12: 6.
- Mullin, Joe. 2013. "Google Books Ruled Legal in Massive Win for Fair Use (Updated)." *Ars Technica*, November 14.
- Murray, Abraham. 2010. "Discover More Than 3 Million Google Ebooks from Your Choice of Booksellers and Devices." *Inside Google Books* (blog). December 6. <http://booksearch.blogspot.com/2010/12/discover-more-than-3-million-google.html>. Accessed July 8, 2013.

- Nadella, Satya. 2008. "Book Search Winding Down." *Bing Community* (blog). May 23.
<http://blogs.msdn.com/livesearch/archive/2006/12/05/live-search-books-beta-release.aspx>.
- Nasaw, David. 2006. *Andrew Carnegie*. New York: Penguin.
- Neubert, Michael. 2008. "'Books.Google.Ru' -- Google's Mass Digitization of Russian-Language Books." *Slavic & East European Information Resources* 9, no. 1: 53-62.
- New York Public Library. 2012. "History of the New York Public Library."
<http://www.nypl.org/help/about-nypl/history>. Accessed July 19, 2012.
- — —. n.d. "Google Books at NYPL." NYPL Help Center.
http://catalog.nypl.org/screens/help_googlebooks_about.html. Accessed April 27, 2011.
- Nielsen, Erland Kolding. 2008. "Digitisation of Library Material in Europe: Problems, Obstacles and Perspectives Anno 2007." *Liber Quarterly: The Journal of European Research Libraries* 18, no. 1: 20-27.
- Notess, Greg R. 2006. "Search Engine Update." *Online* 30, no. 1 (January/February): 15.
- — —. 2007. "Search Engine Update." *Online* 31, no. 2 (March/April): 14.
- Nuendorf, Kimberly A. 2002. *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage.
- Nunberg, Geoffrey. 2009a. "Google Books: A Metadata Train Wreck." *Language Log*, August 29.
- — —. 2009b. "Google's Book Search: A Disaster for Scholars." *Chronicle of Higher Education* (31 August).
- O'Day, Vicki L., and Bonnie A. Nardi. 2003. "An Ecological Perspective on Digital Libraries." In *Digital Library Use: Social Practice in Design and Evaluation*, edited by Ann Peterson Bishop, Nancy A. Van House and Barbara P. Battenfield. 65-82. Cambridge, MA: MIT Press.
- O'Leary, Mick. 2009. "Open Content Alliance Embodies Open Source Movement." *Information Today* 26, no. 1 (January): 37-43.
- Oder, Norman. 2010. "Google Book Search by the Numbers." *Library Journal* (12 February).
- Oder, Norman, Lynn Blumenstein, and Josh Hadro. 2009. "Newsdesk." *Library Journal* 134, no. 20 (December 15): 14-17.
- Oldenburg, Ray. 1991. *The Great Good Place*. New York: Marlowe & Company.
- Open Content Alliance. 2009. "Open Content Alliance: FAQ."
<http://www.opencontentalliance.org/faq>. Accessed October 4, 2009.
- — —. n.d.-a. "About." <http://www.opencontentalliance.org/about/>. Accessed January 22, 2012.
- — —. n.d.-b. "FAQ." <http://www.opencontentalliance.org/faq/>. Accessed January 22, 2012.
- Orlikowski, Wanda J. 1992. "The Duality of Technology: Rethinking the Concept of Technology in Organizations." *Organization Science* 3, no. 3: 398-472.
- Orlikowski, Wanda J., and Stephen R. Barley. 2001. "Technology and Institutions: What Can Research on Information Technology and Research on Organizations Learn from Each Other?" *MIS Quarterly* 25, no. 2 (June): 145-65.
- Orwant, Jon. 2010a. "Find out What's in a Word, or Five, with the Google Books Ngram Viewer." *Inside Google Books* (blog). December 16.
<http://booksearch.blogspot.com/2010/12/find-out-whats-in-word-or-five-with.html>. Accessed July 8, 2013.
- — —. 2010b. "Our Commitment to the Digital Humanities." Official Google Blog.
<http://googleblog.blogspot.com/2010/07/our-commitment-to-digital-humanities.html>. Accessed August 30, 2010.
- Otake, Tomoko. 2009. "Google Books Leaves Japan in Legal Limbo." *Japan Times*, 22 July.

- P1. Interviewed by Elisabeth A. Jones, September 6, 2011. In person.
- P2. Interviewed by Elisabeth A. Jones, September 14, 2011. In person.
- P3. Interviewed by Elisabeth A. Jones, September 14, 2011. In person.
- P4. Interviewed by Elisabeth A. Jones, September 14 & 21, 2011. In person; follow-up on Skype.
- P5. Interviewed by Elisabeth A. Jones, September 9, 2011. In person.
- P7. Interviewed by Elisabeth A. Jones, September 9, 2011. In person.
- P8. Interviewed by Elisabeth A. Jones, October 5, 2011; follow-up conversation June 10, 2013. In person.
- P9. Interviewed by Elisabeth A. Jones, October 4, 2011. In person.
- P10. Interviewed by Elisabeth A. Jones, October 5, 2011. In person.
- P11. Interviewed by Elisabeth A. Jones, October 27, 2011. Skype.
- P12. Interviewed by Elisabeth A. Jones, November 3, 2011. Skype.
- P13. Interviewed by Elisabeth A. Jones, October 26, 2011. Skype.
- P14. Interviewed by Elisabeth A. Jones, January 12, 2012. Skype.
- P15. Interviewed by Elisabeth A. Jones, February 16, 2012. Skype.
- P16. Interviewed by Elisabeth A. Jones, February 15, 2012. Skype.
- P17. Interviewed by Elisabeth A. Jones, January 12 & 24, 2012. Skype.
- P18. Interviewed by Elisabeth A. Jones, January 16, 2012. Skype.
- Page, Larry, and Sergey Brin. 2004. "'An Owner's Manual' for Google's Shareholders." 2004 Founder's IPO Letter. <http://investor.google.com/corporate/2004/ipo-founders-letter.html>. Accessed June 26, 2013.
- Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. "The PageRank Citation Ranking: Bringing Order to the Web." Stanford.
- Patton, Michael Quinn. 2002. *Qualitative Research & Evaluation Methods*. 3rd ed. Thousand Oaks, CA: Sage.
- Peek, Robin. 2007. "2007 Will Be More Open." *Information Today* 24, no. 1: 15-16.
- "People." 2008. *Technicalities* 28, no. 6: 23.
- Peterson, Jennifer. 2008. "New National Poll Shows Library Card Registration Reaches Historic High." American Library Association. <http://www.ala.org/news/news/pressreleases2008/September2008/ORSharris>. Accessed July 20, 2012.
- Petrou, David, Chiu-Ki Chan, Daniel Loreto, Jeffrey C. Reynar, and Nikola Jevtic. 2010. Query-Independent Entity Importance in Books. US Patent 20100281034 A1, issued November 4, 2010.
- Pinch, Trevor, and Wiebe Bijker. 1984. "The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other." *Social Studies of Science* 14, no. 3 (August): 399-441.
- Pon, Cheryl. 2010. "Dubliners." *Inside Google Books* (blog). October 28. http://booksearch.blogspot.com/2010_10_01_archive.html. Accessed July 8, 2013.
- Poole, William Frederick. 1881. *The Construction of Library Buildings*. Circulars of Information of the Bureau of Education. Circulars of Information of the Bureau of Education. Washington: Government Printing Office.
- Posner, Ernst. 1940. "Some Aspects of Archival Development since the French Revolution." *The American Archivist* 3, no. 3: 159-72.
- Powell, Christina Kelleher. 2008. "Opac Integration in the Era of Mass Digitization: The Mbooks Experience." *Library Hi Tech News* 26, no. 1: 24-32.

- Pozzebon, Marlei, and Alain Pinsonneault. 2005. "Challenges in Conducting Empirical Work Using Structuration Theory: Learning from IT Research." *Organization Studies* 26, no. 9: 1353-76.
- Prelinger, Rick. 2006. "Google Book Search Treats Government Documents as Copyrighted Material." *Prelinger Library Blog* (blog). February 18. <http://prelingerlibrary.blogspot.com/2006/02/google-book-search-treats-government.html>. Accessed July 22, 2013.
- President's Information Technology Advisory Committee. 2001. "Digital Libraries: Universal Access to Human Knowledge." Arlington, VA: National Coordination Office for Information Technology Research & Development. February.
- Princeton University Library. 2010. "Google Books Library Project." <http://library.princeton.edu/google/>. Accessed January 17, 2012.
- Proceedings of the Trustees*. 1852-1858. Vol. 1. Boston: Public Library of the City of Boston. Courtesy Boston Public Library: Rare Books & Manuscripts.
- Project Gutenberg. 2008. "About - Gutenberg." <http://www.gutenberg.org/wiki/Gutenberg>About>. Accessed January 26, 2011.
- — —. 2013. "Project Gutenberg - Home." http://www.gutenberg.org/wiki/Main_Page. Accessed December 15, 2013.
- "Pulp Friction." 2005. *Economist* 377, no. 8452: 63-64.
- Quint, Barbara. 2004. "Google and Research Libraries Launch Massive Digitization Project." *Information Today* (20 December).
- — —. 2006a. "Google Book Search Adds Big, Brave Partner: The University of California." *Information Today* (14 August).
- — —. 2006b. "Google Opens Public Domain Books for Downloading, Michigan Launches Mbooks." *Information Today* (5 September).
- — —. 2007. "Who the Heck Is Tristram Shandy? Or What's Not Wrong with Google Book Search?". *Information Today* 24, no. 9 (October): 7-8.
- — —. 2008. "1923-1963: Google Book Search Targeting More Books for Public Domain?". *Information Today* (26 June).
- Raber, Douglas. 1997. *Librarianship and Legitimacy: The Ideology of the Public Library Inquiry*. Westport, CT: Greenwood Press.
- Rapp, David. 2010. "Infotech." *Library Journal* 135, no. 13 (August 1): 16-n/a.
- — —. 2011a. "Infotech." *Library Journal* 136, no. 6 (April 1): 16.
- — —. 2011b. "Infotech." *Library Journal* 136, no. 12 (July 1): 18.
- — —. 2011c. "Infotech." *Library Journal* 136, no. 10 (June 1): 20.
- Read, Brock. 2007. "Some Publishers Warm to Google Book Search." *The Wired Campus* (blog). June 5. <http://chronicle.com/wiredcampus/article/2122/some-publishers-warm-to-google-book-search>.
- Reddy, Raj. 2006. "The Million Book Digital Library Project: Research Problems in Data Mining and Discovery." *Proceedings of the Fifth IEEE International Conference on Data Mining*.
- Redmann, Gail. 1993. "Archivists and Genealogists: The Trend toward Peaceful Coexistence." *Archival Issues* 18, no. 2: 121-32.
- Reinstein, Arielle. 2007. "Thoughts from a Roads Scholar." *Inside Google Books* (blog). March 20. <http://booksearch.blogspot.com/2007/03/thoughts-from-roads-scholar.html>. Accessed July 8, 2013.
- Rejected Settlement Opinion*, Case No. 05 CV 8136-DC The Authors Guild, Inc., Association of American Publishers, Inc., et al. vs. Google, Inc. (2011).

- "Report of the Committee to Examine the Library." 1853. City Document No. 73. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. November 10. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Examining Committee." 1854. City Document No. 74. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. October 31. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Examining Committee." 1856. City Document No. 61. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. November 5. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Examining Committee." 1857. City Document No. 71. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. November 7. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Examining Committee." 1858. City Document No. 46. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. November 4. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Examining Committee." 1859. City Document No. 66. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. November 4. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Examining Committee." 1863. City Document No. 97. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. November 4. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Examining Committee." 1864. City Document No. 92. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. November 11. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Examining Committee." 1865. City Document No. 96. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. November 10. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Examining Committee." 1866. City Document No. 110. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. November 10. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Examining Committee." 1867. City Document No. 114. In *Reports of the Examining Committees, 1853-68*. Boston Public Library. November 11. Courtesy Boston Public Library: Delivery Desk.
- "Report of the Trustees of the Public Library of the City of Boston." 1852. City Document No. 37. Boston: Boston City Council. July 6.
- "Reports of the Examining Committees." 1853-68. Boston Public Library. Courtesy Boston Public Library: Delivery Desk.
- Rice, Prudence M. 2004. *Maya Political Science: Time, Astronomy, and the Cosmos*. Austin, TX: University of Texas Press.
- Rieger, Oya Y. 2008. "Preservation in the Age of Large-Scale Digitization." Washington, DC: Council on Library and Information Resources. February.
- Rieser, Andrew Chamberlin. 2003. *The Chautauqua Moment : Protestants, Progressives, and the Culture of Modern Liberalism*. New York: Columbia University Press.
- "RLG Joins Open Content Alliance." 2005. *College & Research Libraries News* 66, no. 11 (December): 770.
- Rosenblatt, Bill. 2005. "Rights Management and the Revolution in E-Publishing." *INDICARE Monitor* 2, no. 9.

- Roush, Wade. 2005. "The Infinite Library: Does Google's Plan to Digitize Millions of Print Books Spell the Death of Libraries, or Their Rebirth?". *Technology Review* (May): 54-59.
- Samuelson, Pamela. 2009. "Legally Speaking: The Dead Souls of the Google Book Search Settlement." *Communications of the ACM* 52, no. 7 (July): 28-30.
- — —. 2010. "Google Book Search and the Future of Books in Cyberspace." *Minnesota Law Review* 94, no. 5: 1308-74.
- Sandler, Mark. 2005. "Disruptive Beneficence: The Google Print Program and the Future of Libraries." *Internet Reference Services Quarterly* 10, no. 3-4: 5-22.
- Sands, Ryan. 2006. "From the Mail Bag: Public Domain Books and Downloads." *Google Book Search Blog* (blog). November 9. <http://booksearch.blogspot.com/2006/11/from-mail-bag-public-domain-books-and.html>. Accessed July 23, 2013.
- — —. 2009. "From the Mailbag: Blending Books When You Want Them." *Inside Google Books* (blog). March 5. <http://booksearch.blogspot.com/2009/03/from-mailbag-blending-books-when-you.html>. Accessed July 8, 2013.
- Sayer, Peter. 2006. "Update: French Publishers Join Google Booksearch Lawsuit." *InfoWorld* (31 October).
- Schatz, Bruce, and Hsinchun Chen. 1996. "Building Large-Scale Digital Libraries." *Computer* 29, no. 5: 22-26.
- — —. 1999. "Digital Libraries: Technological Advances and Social Impacts." *Computer* 32, no. 2 (February): 45-50.
- Schiff, Jennifer. 2006. "The Wayback Machine: From Petabytes to Petaboxes." *Enterprise Storage Forum*, September 20.
- Schilit, Bill N., and Okan Kolak. 2008. "Exploring a Digital Library through Key Ideas." *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*.
- Schneider, Karen. 2008. "Hurry up, Please. It's Time - State of Emergency." *Serials Librarian* 54, no. 1/2: 57-62.
- Schroeder, Pat, and Bob Barr. 2005. "Reining in Google." *The Washington Times*, November 3.
- Schutt, Russel K. 2009. *Investigating the Social World: The Process and Practice of Research*. Thousand Oaks, CA: Sage.
- Seattle Public Library. 2009. "Rules of Conduct." http://www.spl.org/default.asp?pageID=about_policies_conduct. Accessed August 11, 2010.
- Seaver, Benjamin. 1852. "Message of the Mayor on the Subject of a Public Library." City Document No. 10. City of Boston. Courtesy Boston Public Library: Government Documents Department.
- "Second Annual Report of the Board of Trustees of the Carnegie Library of Atlanta." 1901.
- "Second Annual Report of the Trustees of the City Library." 1854. City Document No. 74. Boston: City of Boston. November 2.
- "Settlement FAQ." 2008. SU LAIR, Stanford University. <http://lib.stanford.edu/google-books/settlement-faq>. Accessed April 27, 2011.
- "Seventh Annual Report of the Trustees of the City Library." 1859. City Document No. 66. Boston: City of Boston. November 10.
- Shapiro, Gilbert, and John Markoff. 1997. "A Matter of Definition." *Text Analysis for the Social Sciences*.
- Shera, Jesse H. 1949. *Foundations of the Public Library*. Chicago: University of Chicago Press.
- — —. 1952. "On the Value of Library History." *The Library Quarterly* 22, no. 3: 240-51.
- Shin, Dong-Hee. 2006. "Effective Design in the Development of Public Information Infrastructure: A Social Constructionist Approach." *Information Polity* 11: 85-100.

- Sismondo, Sergio. 1993. "Some Social Constructions." *Social Studies of Science* 23, no. 3 (August): 515-53.
- "Sixth Annual Report of the Trustees of the City Library." 1858. City Document No. 46. Boston: City of Boston. November 15.
- Slawski, Bill. 2011. "Searching in Google's Book Search (the Seo of Books?)." *SEO by the Sea: Getting Information about Search and SEO Directly from the Search Engines* (blog). June 7. <http://www.seobythesea.com/2011/06/searching-in-googles-book-search-the-seo-of-books/>. Accessed August 3, 2013.
- "Sloan Foundation Grant Awarded." 2006. *Opencontentalliance.org* (blog). December 19. <http://www.opencontentalliance.org/2006/12/19/sloan-foundation-grant-awarded/>. Accessed February 2, 2012.
- Small, Herbert. 1895. *Handbook of the New Public Library in Boston*. Boston: Curtis and Company.
- Smith, Adam. 2005. "Making Books Easier to Find." *Google Blog* (blog). August 11. <http://googleblog.blogspot.com/2005/08/making-books-easier-to-find.html>.
- Smith, Kevin. 2013. "Parallel Tracks, Parallel Successes." *Scholarly Communications @ Duke* (blog). July 2. <http://blogs.library.duke.edu/scholcomm/2013/07/02/parallel-tracks-parallel-successes/>. Accessed July 5, 2013.
- Sporkin, Andi. 2012a. "Publishers and Google Reach Settlement." Association of American Publishers. <http://www.publishers.org/press/85/>. Accessed October 11, 2012.
- — —. 2012b. "Publishers and Google Reach Settlement." Association of American Publishers. <http://www.publishers.org/press/85/>. Accessed July 5, 2013.
- St. Clair, Gloriana. 2005. "Million Book Project Vs. Google Print." *Journal of Zhejiang University (Science)* 6A, no. 11: 1195-200.
- — —. 2008. "The Million Book Project in Relation to Google." *Journal of Library Administration* 47, no. 1/2: 151-63.
- "Staff Bios." n.d. Internet Archive. <http://archive.org/about/bios.php>. Accessed August 21, 2013.
- "Stanford and Google Book Search: Statement of Support and Participation." 2005. SU LAIR, Stanford University. <http://lib.stanford.edu/google-books/statement-support-participation-2005>. Accessed April 27, 2011.
- Stanford University Libraries. 2006. "Stanford Google Library Project: FAQ." <https://lib.stanford.edu/google-books/faq>. Accessed January 17, 2012.
- "Stanford's Role in Google Books." 2010. SU LAIR, Stanford University. <http://lib.stanford.edu/google-books/stanfords-role-google-books>. Accessed April 27, 2011.
- State of Massachusetts. 1848. "An Act to Authorize the City of Boston to Establish a Public Library."
- Stevenson, William M. 1892. "Second Annual Report of the Carnegie Free Library, Allegheny, PA, for the Year Ending February 29th, 1892." Allegheny, PA: Press of John Ogden. February 29.
- Stone, Elizabeth W. 1977. *American Library Development 1600-1899*. New York: The H.W. Wilson Company.
- Strauss, Anselm, and Juliet M. Corbin. 1997. *Grounded Theory in Practice*. Thousand Oaks, CA: Sage.
- Stripling, Jack. 2008. "Still Searching." *Inside Higher Ed* (29 October).
- Strum, Charles. 1992. "Belleville Journal; Restoring Heritage and Raising Hopes for Future." *New York Times*, March 2.
- Suber, Peter. 2005. "The Open Content Alliance." *SPARC Open Access Newsletter*, November 2.

- — —. 2006. "Why Google Blocks Access to Public Domain Books Outside the US." *Open Access News* (blog). November 10. <http://www.earlham.edu/~peters/fos/2006/11/why-google-blocks-access-to-public.html>. Accessed July 23, 2013.
- — —. 2009. "Open Access and the Google Book Settlement." *SPARC Open Access Newsletter*, no. 140.
- Sullivan, Danny. 2007. "Authorama: Testing If Google Can Restrict Public Domain Books It Offers for Download." *Search Engine Land*, January 10.
- Swartz, Aaron. 2007. "Announcing the Open Library." *Raw Thought* (blog). July 16. <http://www.aaronsw.com/weblog/openlibrary>. Accessed August 21, 2013.
- Taycher, Leonid. 2010. "Books of the World, Stand up and Be Counted! All 129,864,880 of You." *Inside Google Books*, August 5.
- "Technology." n.d. Alexa Internet. <http://www.alexa.com/company/technology>. Accessed August 21, 2013.
- "Third Annual Report of the Librarian of Carnegie Free Library, Allegheny, PA, for the Year Ending 28th February, 1893." 1893. February 28.
- "Third Annual Report of the Trustees of the City Library." 1855. City Document No. 57. Boston: City of Boston. November 15.
- "Thirteenth Annual Report of the Trustees of the City Library." 1865. City Document No. 96. Boston: City of Boston. November 27.
- Thompson, Charles Seymour. 1952. *Evolution of the American Public Library*. Metuchen, N.J.: Scarecrow Press.
- Tong, Judy. 2002. "Responsible Party - Brewster Kahle; a Library of the Web, on the Web." *New York Times*, September 8.
- Toobin, Jeffrey. 2007. "Google's Moon Shot: The Quest for the Universal Library." *The New Yorker*, February 5.
- Townsend, Robert B. 2007. "Google Books: Is It Good for History?." *Perspectives on History* 45: 45-47.
- "The Triangle Research Libraries Network (TRLN) Member Libraries Join Open Content Alliance." 2008. *D-Lib Magazine* 14, no. 3/4.
- "TRLN Libraries Join Open Content Alliance." 2008. *Advanced Technology Libraries*.
- "TRLN Member Libraries Join Open Content Alliance." 2008. *Library Hi Tech News* 25, no. 4: 21-21.
- Trustees of the Public Library. 1865. *Memorial of Joshua Bates*. Boston. Courtesy Boston Public Library: Rare Books & Manuscripts.
- "Twentieth Annual Report of Carnegie Free Library and Sixteenth Annual Report of Carnegie Club, Braddock, PA, for the Year 1908, Together with Statistics from the Founding of the Library." 1908. Braddock, PA.
- Tyack, David B. 1967. *George Ticknor and the Boston Brahmins*. Cambridge, MA: Harvard University Press.
- "U of Illinois Joins Open Content Alliance." 2007. *Advanced Technology Libraries* 36, no. 4.
- "UNC-Chapel Hill Library and Library School Join Open Content Alliance." 2006. *College & Research Libraries News* 67, no. 3 (March): 140.
- Universiteitsbibliotheek Gent. 2007. "Google Project: In a Nutshell." <http://lib.ugent.be/info/en/project-google.shtml>. Accessed April 27, 2011.
- University of California, and Google, Inc. 2006. "Cooperative Agreement." August 3.
- University of California Libraries. 2005. "UC Libraries Partner with Technology Companies and Non-Profits to Provide Free Access to Digitized Books." October 3.

- University of Michigan. 2004a. "Google/U-M Project Opens the Way to Universal Access to Information." December.
- — —. 2004b. "U-M University Library – Google Digitization Project: A Brief Overview."
- — —. 2005. "UM Library/Google Digitization Partnership FAQ, August 2005." August.
- — —. n.d.-a. "Michigan Digitization Project - Google Settlement Agreement FAQ." <http://www.lib.umich.edu/michigan-digitization-project/michigan-digitization-project-google-settlement-agreement-faq>. Accessed January 20, 2012.
- — —. n.d.-b. "Settlement Agreement between Google and Plaintiffs the Authors Guild and the Association of American Publishers." http://www.vpcomm.umich.edu/pa/key/google_settle.html. Accessed January 20, 2012.
- University of Michigan, and Google, Inc. 2004. *Cooperative Agreement*.
- — —. 2009. "Amendment to Cooperative Agreement." May 19.
- University of Michigan Libraries. 2008. "Michigan Digitization Project Workflow."
- — —. 2011. "Michigan Digitization Project." <http://www.lib.umich.edu/michigan-digitization-project>. Accessed January 17, 2012.
- University of Texas, and Google, Inc. 2007. "Cooperative Agreement." January 8.
- — —. 2009. "Amendment to Cooperative Agreement." June 8.
- University of Texas Libraries. n.d. "FAQs About the University of Texas - Google Books Library Project." <http://www.lib.utexas.edu/google/faqs.html>. Accessed January 17, 2012.
- University of Virginia, and Google, Inc. 2010. "Amended Cooperative Agreement."
- University of Virginia Library. 2006. "Google Book Library Project: University of Virginia Joins Leading Research Libraries in Partnership with Google to Increase Discovery of Knowledge – and to Offer Library Books to Global Audience." November 14.
- — —. n.d.-a. "About the U.Va./Google Book Library Project." <http://www2.lib.virginia.edu/press/uvagoogole/about.html>. Accessed January 17, 2012.
- — —. n.d.-b. "Frequently Asked Questions (FAQs) About the University of Virginia/Google Book Library Project." <http://www2.lib.virginia.edu/press/uvagoogole/faq.html>. Accessed January 17, 2012.
- University of Wisconsin. 2006. "UW-Madison Joins Massive Google Book Project." October 12.
- University of Wisconsin, and Google, Inc. 2006. "Cooperative Agreement."
- — —. 2009. "Amendment to Cooperative Agreement."
- University of Wisconsin Libraries. n.d.-a. "University of Wisconsin - Madison Google Digitization Initiative: FAQs." <http://www.library.wisc.edu/digitization/faq.html#primary>. Accessed January 20, 2012.
- — —. n.d.-b. "University of Wisconsin - Madison Google Digitization Initiative: Principles and Values." <http://www.library.wisc.edu/digitization/#about>. Accessed January 20, 2012.
- Vaidhyathan, Siva. 2005. "A Risky Gamble with Google." *Chronicle of Higher Education*, December 2.
- — —. 2007. "The Googlization of Everything and the Future of Copyright." *University of California, Davis Law Review* 40: 1207-31.
- — —. 2011. *The Googlization of Everything (and Why We Should Worry)*. Berkeley: University of California Press.
- Vaidhyathan, Siva, and Brewster Kahle. 2011. *Googlization and the Future of Books*. San Francisco: Commonwealth Club of California. Podcast. Recorded May 17, 2011.

- Accessed via: <http://www.commonwealthclub.org/events/2011-05-17/googlization-and-future-books>
- van Dijk, A. 2005. "Primary Source Publishers in a Googling World." *Microform & Imaging Review* 34, no. 1: 31-34.
- Van Gemert, Ed. 2009. "UW-Madison Expands Agreement with Google." July 9.
- Van House, Nancy A., Ann Peterson Bishop, and Barbara P. Battenfield. 2003. "Introduction: Digital Libraries as Sociotechnical Systems." In *Digital Library Use: Social Practice in Design and Evaluation*, edited by Ann Peterson Bishop, Nancy A. Van House and Barbara P. Battenfield. 1-21. Cambridge, MA: MIT Press.
- Van Slyck, Abigail A. 1991. "'The Utmost Amount of Effectiv [Sic] Accommodation': Andrew Carnegie and the Reform of the American Library." *The Journal of the Society of Architectural Historians* 50, no. 4 (December): 359-83.
- — —. 1995. *Free to All: Carnegie Libraries & American Culture, 1890-1920*. Chicago: University of Chicago Press.
- — —. 1996. "The Lady and the Library Loafer: Gender and Public Space in Victorian America." *Winterthur Portfolio* 31, no. 4: 221-42.
- — —. 2001. "The Librarian and the Library: Why Place Matters." *Libraries & Culture* 36, no. 4: 518-23.
- Verba, Sidney. 2004. "A Letter from Sidney Verba." Harvard University Library News Archive. http://hul.harvard.edu/news/2004_1213_letter.html. Accessed April 27, 2011.
- Vincent, Luc. 2008. "Google Book Search: Document Understanding on a Massive Scale." Paper presented at *9th International Conference on Document Analysis and Recognition*, Piscataway, NJ. IEEE.
- Wadlin, Horace D. 1911. *The Public Library of the City of Boston: A History*. Boston: Trustees of the Public Library.
- Wang, Xiaoyue, Lexiang Ye, Eamonn Keogh, and Christian Shelton. 2008. "Annotating Historical Archives of Images." Paper presented at *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, Pittsburgh PA. ACM.
- Weisl, Annabella. 2010. "Unlocking Our Shared Cultural Heritage: Google Partners with the Austrian National Library." *Google Europe Blog* (blog). June 15. http://googlepolicyeuropa.blogspot.com/2010/06/unlocking-our-shared-cultural-heritage_4958.html. Accessed July 8, 2013.
- — —. 2011. "Printed Treasures from the Golden City." *Inside Google Books* (blog). February 28. <http://booksearch.blogspot.com/2011/02/printed-treasures-from-golden-city.html>. Accessed July 8, 2013.
- West, Jessamyn. 2011. *Without a Net: Librarians Bridging the Digital Divide*. Santa Barbara, CA: Libraries Unlimited.
- Wheeler, Joseph L., and Alfred Morton Githens. 1941. *The American Public Library Building: Its Planning and Design with Special Reference to Its Administration and Service*. Chicago: American Library Association.
- Whitehill, Walter Muir. 1956. *Boston Public Library: A Centennial History*. Cambridge, MA: Harvard University Press.
- Wiegand, Wayne A. 1999. "Tunnel Vision and Blind Spots: What the Past Tells Us About the Present; Reflections on the Twentieth-Century History of American Librarianship." *The Library Quarterly* 69, no. 1: 1-32.
- — —. 2000. "American Library History Literature, 1947-1997: Theoretical Perspectives?" *Libraries & Culture* 35, no. 1: 4-34.

- Wilkin, John. 2008. "The Launch of HathiTrust." *John Wilkin's Blog* (blog). October 13. Accessed October 13, 2008.
- Wilkin, John P. 2011. "Bibliographic Indeterminacy and the Scale of Problems and Opportunities of 'Rights' in Digital Collection Building." *Ruminations* (February).
- Willett, Perry. 2009. "Mass Digitization and Its Impact on Interlending and Document Supply." *Interlending & Document Supply* 37, no. 3: 143-48.
- Williams, Patrick. 1988. *The American Public Library and the Problem of Purpose*. Westport, CT: Greenwood Press.
- Williamson, Charles C. 1923. *Training for Library Service: A Report Prepared for the Carnegie Corporation of New York*. Boston: The Merrymount Press.
- Wilson, Andrew Norman. 2009-2011. "Workers Leaving the Googleplex." [HD Video]. <http://www.andrewnormanwilson.com/WorkersGoogleplex.html>. Accessed July 18, 2013.
- Winsor, Justin. 1871. "The Boston Public Library." *Scribner's Monthly* 3, no. 2 (December): 150-56.
- Winthrop, Robert Charles, Mayor Jerome V.C. Smith, Rev. E.N. Kirk, and others. 1855. "Proceedings on the Occasion of Laying the Corner Stone of the Public Library." Boston: Moore & Crosby, City Printers. September 17.
- Wittenborg, Karin. 2009. Letter in support of final settlement, The Author's Guild et al v. Google, to Denny Chin, September 3 http://www2.lib.virginia.edu/press/uvagoogole/uva_librarian_letter_of_support.pdf.
- Wray, Richard. 2006. "Legal Victory for Google in Library Project." *Guardian*, 29 June.
- Yahoo. 2005. "Global Consortium Forms Open Content Alliance to Bring Additional Content Online and Make It Searchable." October 3.
- "Yahoo Takes on Google." 2005. *Bookseller*, October 7, 8.
- Yin, Robert K. 2003. "Case Study Research: Design and Methods." *Applied Social Research Methods Series* 5.
- Young, Jeffrey R. 2005a. "From Gutenberg to Google: Five Views on the Search-Engine Company's Project to Digitize Library Books." *Chronicle of Higher Education* (3 June): A24.
- — —. 2005b. "Microsoft Joins Digital-Library Consortium." *Chronicle of Higher Education* 52, no. 12: A35.
- — —. 2006a. "Scribes of the Digital Era." *Chronicle of Higher Education* 52, no. 21: A34-A36.
- — —. 2006b. "U. Of Michigan Puts Digitized Books Online but Bars Full Text of Copyrighted Works." *Chronicle of Higher Education* 53, no. 3 (September 8): A32.
- Zhao, Lisa. 2006. "How Will Google Print Library Project Affect the Use of Books?". *Journal of Library and Information Science* 32, no. 1 (April): 18-24.
- Zickuhr, Kathryn. 2013. "Who's Not Online and Why." Internet and American Life Project. Washington, DC: Pew Research Center. September 25.