Open Data for Public Good

Grant RE-40-16-0015-16

Final Report Narrative

Performance period:
July 1, 2016 - March 31, 2021

Dr. Carole Palmer, Co-PI
clpalmer@uw.edu

Dr. Nicholas Weber, Co-PI
nmweber@uw.edu

Bree Norlander, Project Coordinator
norlab@uw.edu

Kaitlin Throgmorton, Research Assistant
kaitlin@kaitlinthrogmorton.com

Lindsey Schwartz, Research Assistant
lsschwar@uw.edu

Stacey Wedlake, Project Evaluator
staceyaw@uw.edu

University of Washington, Information School

Table of Contents

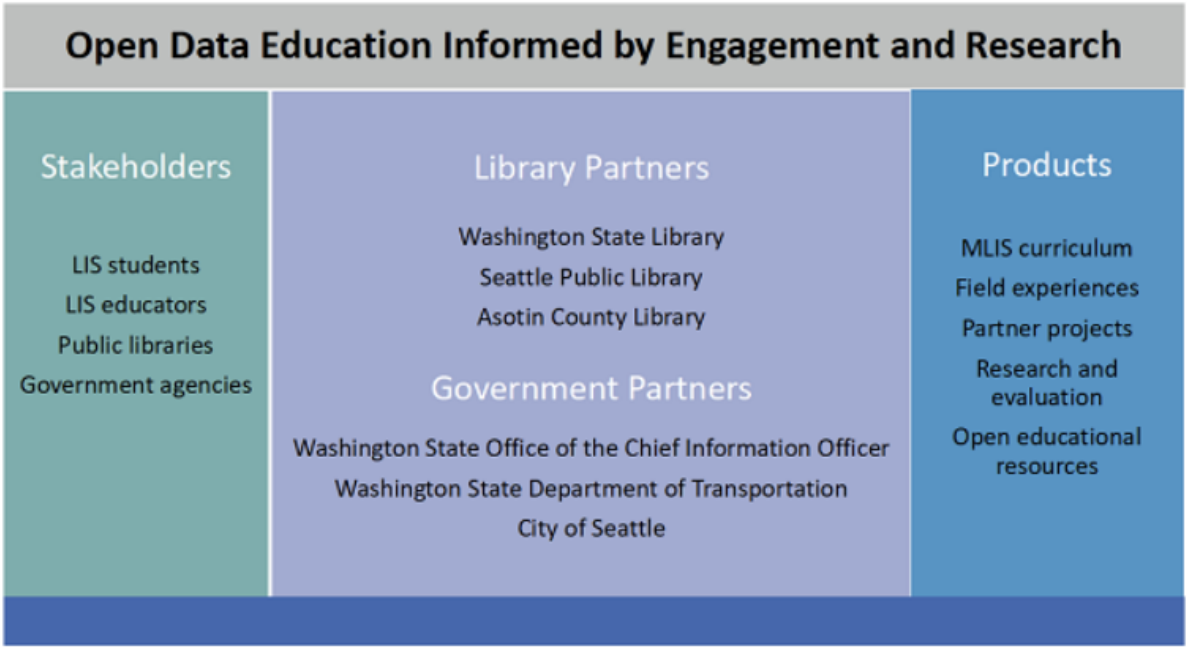
Overview

Introduction

The Open Data for Public Good (ODPG) project began July 1, 2016. The purpose of ODPG's work is to:

1. Prepare future and current public librarians and information professionals to
    a. curate collections of open data of value to local communities
    b. build the necessary infrastructure and preservation environments to sustain open data collections for long-term sustainability of these valued assets
2. Collaborate with public sector agencies and open data providers on advocacy and outreach activities that increase awareness about and use of open data by the public.

The approach, including stakeholders, partners, and products, is outlined in the figure below.



Before ODL began, few open data literacy initiatives existed in the field of Library and Information Science (LIS). At that point in 2016, open data had been encouraged in both government and scientific communities for more than a decade, but library and information science (LIS) professionals had not yet embraced the potential of the open data movement for the general public. ODL aimed to bring current knowledge into practice, particularly for public libraries and public sector agencies, by educating information professionals to harness the open data movement and promote this new, important kind of information resource for the communities they work with and serve. Due to an evolving focus on education and improving open data competencies more generally within the public sector, the Open Data for Public Good

(ODPG) project came to be known as the Open Data Literacy (ODL) project, as reflected in the project website (https://odl.ischool.uw.edu) and other project outputs, including this report.

Over its five-year trajectory, the ODPG/ODL project had three broad areas of focus: 1) curriculum, 2) fieldwork, and 3) research. This multi-faceted approach formed a strong foundation for educating new professionals but also for supporting and promoting the uptake of open data work by libraries and the broader public sector. Partnerships were fundamental to building capacity and assuring practical applicability of our interdependent education, engagement, and research activities. Library partners included urban, rural, and state libraries; government partners spanned both state and city agencies. The integrated, collaborative approach succeeded at creating vital connections between students and practitioners, and between theory and practice, with tangible outcomes, particularly new curriculum and field experiences for students, but also publications, resources, and job placements for students.

Toward the close of the grant period, we invited Stacey Wedlake, a researcher from the Technology and Social Change group in the Information School, to conduct a formal, summative evaluation. Her report extends the formative evaluations conducted throughout the project by providing a more complete and holistic account of student experiences and outcomes. In particular, her report includes results from a survey that adds perspectives from students who completed ODL coursework and/or an internship. A more detailed summary can be found in the Evaluation section, p. 30 in this report, and the full evaluation is attached as Appendix 1.

Curriculum

To meet ODL's purpose of preparing future library and information professionals to curate open data and build supporting infrastructure, the faculty team members completed significant curriculum revision and development within the Information School at the University of Washington. We began by expanding the existing curation curriculum, previously focused on research data management (RDM) and open science and research practices. An on-ramp to broader open data expertise was produced through a series of three classes: Data Curation I, Data Curation II, and Digital Preservation. Curriculum was enhanced with addition of new modules specifically on open government data but also with increased exposure to hands-on learning and application of data curation technologies and tools and guest lecturers featuring real-world data curation experts and open data activities in practice.

The curriculum has been well-received by students. For example, the entry-level course, Data Curation I, has frequently filled at or above capacity immediately after registration opens. As noted by the MLIS program advisor, courses that fill that quickly are seen by students as critical to one or more career paths. They also noted that more students are indicating interest in data curation and digital preservation in their admissions applications. Students have consistently created impressive portfolios of data curation work through the sequence of courses. Many also participated in the ODL internship program, discussed below, or sought out data curation directed fieldwork opportunities or independent study experiences.

In addition to the graduate coursework, ODL produced open curriculum aimed at library and information professionals on preparing and publishing open data for the general public, using library open data as examples. This curriculum is meant to fill a gap ODL identified in the existing tutorial landscape. Most available public curriculum covers the fundamentals and importance of open data, but doesn't actually provide step-by-step guidance on how to prepare and publish open data for the first time. The team also assessed the ever growing abundance of materials available on basic open data topics, data science skills, and related competencies, to provide a selective set of materials well suited to the learning needs of public librarians and other public sector professionals interested in open data.

Student Fieldwork Experiences

A highly successful internship program was a cornerstone of the ODL project, made possible by the strong relationships with committed external partners. The internships gave graduate students the opportunity to advance their open data expertise by translating theoretical classroom concepts into practical experience in the field.

From 2017-2020, ODL facilitated 12 full-time, 8-week summer internships and two part-time internships (14 total). Interns were matched with public sector organizations based on alignment of student interests and skills with sponsor internal needs. Internship projects were designed and implemented in collaboration with a sponsor representative who served as a mentor throughout the internship. Each year, the cohort of interns presented their projects at an ODL Internship Showcase event. Some students also disseminated their work at conferences and in academic journals. Many interns now work in professional roles that relate to their ODL experience. Most notably, Kathleen Sullivan, who was an ODL intern for Washington State Library (WSL) in 2018, was hired at WSL in a full-time permanent position as open data consultant, a position based on her internship contribution, She later sponsored two additional ODL interns and a 2019-20 Capstone project. See "Selected Alumni Placements" below (p. 16).

In interviews and ODL meetings and events, partners reported overwhelmingly positive views of the internship program. In our final ODL Reunion event, librarians from Asotin County Library explained that partnering with ODL legitimized their open data efforts, and the Washington State Librarian stated that open data was a 'huge win' for reestablishing the importance of public libraries.

Research

Most of ODL's research has focused on public library open data activity, in addition to evaluations of curriculum and fieldwork, particularly student and sponsor perspectives and experiences. Two prominent research projects undertaken during the course of the grant included 1) a systematic evaluation of library data present on open government data portals (Weber & Norlander, 2019) and 2) a survey of public library activity and interest in open data in

the state of Washington, conducted in collaboration with the Washington State Library (Throgmorton et al., 2020). Taken together, these two studies contribute to the field's understanding of the current and aspirational state of library open data work. Results show that public libraries, large and small, are interested in becoming active in open data efforts, and that the amount of library data currently available through public data portals is sparse.

Sustainability

ODL accomplishments will be sustained through long-term curriculum, graduates of the program, and new activities seeded by the grant initiative. Significantly, all three courses were added into the iSchool's annual course offerings in 2021, ensuring the longevity of the open data literacy curriculum after the close of the grant period. This formalization represents a formal contribution to the iSchool's learning objectives for emerging information professionals and responds to continuous positive reception amongst students.

We are also seeing interest in continued investment by partners in sustaining internships and sponsoring capstone projects for students after the grant ends, especially within the Washington State government.

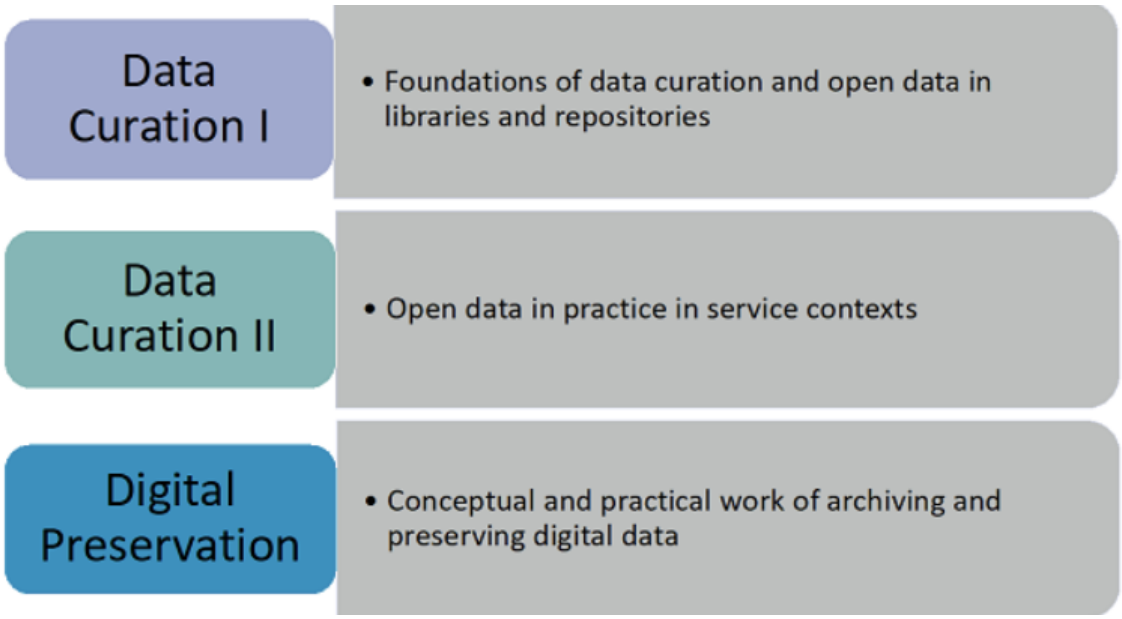
Graduates entering the workforce are a significant part of ODL sustainability. While our tracking is not comprehensive, placement of students who have graduated with ODL coursework and internship experience has been strong. These students will continue to advance open data practices in their institutions. In particular, the professional position, Open Data Literacy Consultant, at the Washington State Library was a direct outcome of the ODL internship program and will continue to foster open data progress in libraries across the state. The partnership established with Washington State agencies and their experience with ODL education has resulted in new Data Steward positions in the Department of Licensing, and investment in new data curriculum developed for the new positions.

The team has also made progress on plans for sustaining other segments of the initiative through the transitional work with the Technology and Social Change Group (TASCHA) at the UW Information School. TASCHA contributors have been essential to our ability to continue coordination with other state libraries. They have also leveraged ideas from one of the outstanding student internship projects for a new IMLS grant funded effort--Leveraging Use of Open Data by Public Library Staff for Community Benefit. The project will develop outreach plans for public libraries by utilizing open data. TASCHA is also committed to extending some ODL activities after the grant period, as part of their Future of Libraries initiative.

Activities

Course Development, Revision, and Enhancement

A three course sequence was developed to prepare students for professional open data work in the public sector.



Fundamentals of Data Curation / Data Curation 1 - LIS 545 (formerly INFX 551)
The first data curation course covers the foundations of data curation and open data in libraries and repositories. Students develop a protocol for curating open data; create new metadata and documentation to make this data accessible; and then publish data to a public repository following W3C's conventions for open data on the web.

Since its debut, the Fundamentals of Data Curation course was revised to include learning objectives related to open data and public sector information, and introductory exposure to relevant technologies and tools. Enrolled students represented the Masters in Library and Information Science (MLIS), Masters in Information Management (MsIM) graduate degree programs, as well as the Informatics undergraduate major in the iSchool.

Beginning in Year 2, we enhanced classroom learning by bringing in professionals with real-world experience and implementing hands-on activities with tools for work with open data. Two new guest lecturers were recruited to represent key areas in open data:

- State level open data and public records efforts
- Current innovations in health data curation

Four technology topics were incorporated into the course, including in-class tutorials listed below. These tutorials are being repurposed for application outside the formal classroom, as professional development resources for practicing public librarians.

- Introduction to GitHub
- Data access through APIs: http://odl.ischool.uw.edu/api_tutorial/
- Data cleaning with Open Refine: http://odl.ischool.uw.edu/openrefine_tutorial/
- Introduction to CKAN and Dataverse open data repository platforms.

The tutorials were developed by Dr. Weber and a PhD student assistant who gained invaluable experience in developing and delivering open data content in the classroom and redesigning the materials for a professional audience.

In year 1, a Data Curation Workbench was developed through collaboration with the National Data Service (NDS), was made available to students in the data curation courses, and all but one in the advanced section used the workbench to demonstrate their final open data project in CKAN It allowed both on-campus and online students to gain experience and experiment with the CKAN and Dataverse software platforms, without the interference of setup and administration. Unfortunately, we were not able to sustain the support of the NDS staff to continue the pilot.

Advanced Data Curation / Data Curation 2 - LIS 546 (formerly LIS 598 or INFX 598)
The second data curation course concentrates on open data in practice in service contexts. Building on skills gained in Data Curation 1, students engage in a series of lab exercises to understand integrating different open data sources, obtain data using a public API, create data models that describe the contents of integrated data, and then engage in a multi-week activity of configuring existing data repository software (CKAN and Dataverse) to publish new collections of integrated open data to the web.

Like the Fundamentals course, Advanced Data Curation was also redesigned to expand the scope to cover open government and civic data, in addition to the previous emphasis on the curation of research data. New content included readings and lectures on smart cities, data privacy, as well as issues around access, value, and quality. Assignments focused on giving students practical experience with curation tools, evaluating open data, and both building and implementing collections designed for a particular user community.

A significant curriculum output is the full documentation of Dr. Weber's Data Curation 2 course: https://nniiicc.github.io/LIS-598-DC2-Sp2020/

Digital Preservation - LIS 506 (formerly LIS 598 or INFX 598)
Digital Preservation consists of conceptual and practical work archiving and preserving digital data. Conceived of as the final course in the ODL sequence, it includes lab exercises that focus on carrying out maintenance tasks for existing digital data collections, and then translating skills in curating data to verify and validate collections of data that are found on the web.

For a list of Curriculum Outputs, see the Outreach and Dissemination section, p. 19.

Student Enrollment
Course enrollment was consistently solid throughout the project performance period, listed below by academic year.

2020-21

Data Curation 1 (two sections)	48
Data Curation 2	11
Digital Preservation	33

2019-20

Data Curation 1 (two sections)	42
Data Curation 2	12
Digital Preservation	33

2018-19

Data Curation 1	34
Data Curation 2	15
Digital Preservation (two sections)	55

2017-18

Data Curation 1 (two sections)	37
Data Curation 2	3
Digital Preservation	23

2016-17

Data Curation 1 (two sections)	35
Data Curation 2	6
Digital Preservation	34

Student Field Experiences

Throughout the project, ODL has worked with project partners to identify and place students in real-world open data settings, through which they gained valuable hands-on experience managing, curating, and visualizing open data in collaboration with a public institution. In total, 14 master's students completed summer internships — 12 full-time and 2 part-time, and 18 students from the MLIS, MSIM, and Informatics programs completed a capstone project.

Internships

Students documented their internship progress via Github, blog posts (https://medium.com/open-data-literacy/archive), and short presentations to the ODL team, partners, and stakeholders in a showcase event at the end of the summer. Projects are outlined below. For a full list of citations of publications, presentations, and blog posts, see the Outreach and Dissemination section (p. 19) below.

Context for Communities: Seattle and Metadata Standards for Civic Data 2017
Student: Nina Showell, Masters of Science in Information Management Program
Sponsor: City of Seattle
Worked with the Open Data Program at the City of Seattle to apply civic metadata standards to the datasets included on data.seattle.gov. There are numerous standards that are currently available, but the way they are applied varies. Data about some areas of civic life, such as bus transit times, corresponds well to existing metadata standards. For others, such as the data generated by cutting-edge Internet of Things devices, standards barely exist. The project researched and reported on the question: How can the City of Seattle reconcile these differing metadata standards and apply them to the city's datasets?
Github | Blog #1 | Blog #2 | Blog #3

OCIO Geospatial Program 2017
Student: M. Wynn Tranfield, Masters of Library and Information Science Program
Sponsor: Washington State Office of the Chief Information Officer
Work in the Geospatial Program Office to evaluate and consolidate Washington State's three main geoportals into one. The multiple authoritative portals were expensive, inefficient, and confusing for new users. Transitioning requires indexing geospatial and non-geospatial data from disparate portals, tracking departmental provenance, and constructing a sustainable portal that can be easily utilized by stakeholders with varying abilities. To that end, the project reviewed the literature on current geoportal best practices and produced an eight-week plan for completion. The completed consolidation and new, singular geoportal can now be found at: https://ocio.wa.gov/geospatial-program-office/geospatial-data.
Github | Blog #1 | Blog #2 | Blog #3

Up and Running with Open Data: Open Data Unconference 2017
Student: Sarah Carrier, Master of Science in Information Management Program
Sponsor: Seattle Public Library
Worked with Seattle Public Library to develop an Open Data Unconference, entitled Up and Running with Open Data, framework. The plan included detailed event logistics, goals, learning objectives, lesson plans, and curriculum. The public *Up and Running with Open Data* event wasSeptember 9, 2017; the regional library staff training was September 20, 2017.
Github | Blog #1 | Blog #2

Washington State Department of Transportation's Data or Term Search Tool 2017
Student: Tim Blankemeyer, Master of Library and Information Science Program
Sponsor: Washington State Department of Transportation (WSDOT)
In collaboration with technical and business stakeholders, the project undertook the initial work for a full business analysis of WSDOT Data or Term Search (DOTS), which is managed by the Data Management Services (DMS) group. At its heart, DOTS is a data catalog: an inventory of all of the data objects within DMS-managed environments, along with technical metadata for those objects. The analysis supported data and domain experts in delivering greater benefit to knowledge seekers through improvements to DOTS and related information-sharing processes.
Github | Blog #1 | Blog #2 | Blog #3

Open Data and the Washington State Library 2018
Student: Kathleen Sullivan, Masters of Library and Information Science Program
Sponsor: Washington State Library
Open government data advocates believe this relatively new information source can improve government transparency and efficiency and provide a platform for innovation, but much of the general public remains unfamiliar with it. The project produced an environmental scan for the Washington State Library to explore two little-researched topics: the extent of open data published by local government, and public libraries' roles in open data publishing and instruction. Based on interviews, recommendations were produced advising that the State Library add a central open data resource guide to the "Library Services" section of its websit and explore options for providing a statewide platform for local open data publishing. The project proved so valuable to the Washington State Library that they initially extended the position as a temporary contract to continue the work for several more months and then transitioned it to a regular professional position of Open Data Literacy Consultant.
Github | Blog #1 | Blog #2

Implementing an Open Data Alliance 2018
Student: Leslie Denning, Masters of Library and Information Science Program
Sponsor: Washington State Office of the Chief Information Officer & The City of Seattle
Researched public record request trends using request logs and interviews. The analysis summarized concerns about the open data/public records relationship and presented possible recommendations including a plan to create an Open Data Alliance. Produced an Open Data Alliance Guidebook to support and formalize a coalition of professionals concerned with Open Data with a specific focus on Public Records Requests as a commonality across jurisdictions.
Github | Blog #1 | Blog #2

Creating an Open Data Alliance 2018
Student: Kevin McCraney, Masters of Science in Information Management Program
Sponsor: Washington State Office of the Chief Information Officer & The City of Seattle
Analyzed a set of public records data requests from across western Washington to assess support for the proactive disclosure of datasets. Requests were categorized using Python and a natural language processing package. Provided recommendations to support implementation.

Github | Blog #1 | Blog #2

Washington State Department of Transportation's Public Disclosure Requests 2018
Student: Angela Gonzalez-Curci, Master of Library and Information Science Program
Sponsor: Washington State Department of Transportation (WSDOT)
In concert with WSDOT moving its data to a new platform, assisted with techniques to identify and proactively release data and documents commonly requested through the Public Disclosure Request process. Produced a ranked list of the most frequently requested types of datasets and records, and assessed the types of exemptions associated with that type of request and the least restricted datasets and records for open disclosure.
Github | Blog #1 | Blog #2

Publishing Library Datasets 2019
Student: Lillian Curanzy, Master of Library and Information Science Program
Sponsor: Asotin County Library
Worked with librarians and government partners as a follow up to activities with the Data Equity for Main Street program (a project that had assistance from a former ODL intern, Kathleen Sullivan). Supported Asotin County Library in their local open data initiative and collaboration with government partners to begin publishing open data. Working over two quarters, completed dataset identification and selection process, gathered and cleaned library datasets, and created an open data policy draft for the county.
Github (private) | Blog #1 | Blog #2 | Blog #3

Data Management Plans for Washington State Department of Transportation 2019
Student: Joan Hua, Master of Library and Information Science Program
Sponsor: Washington State Department of Transportation (WSDOT)
Worked to help WSDOT comply with federal requirements about open access to research data and meeting the business needs of its units. To support creation of data management plans (DMP) and templates, developed a data asset inventory, reviewed data management guidance, and aligned WSDOT to transportation sector best practices for DMP. Reviewed DMPs from other organizations, inventoried data assets, and conducted interviews with subject experts across the organization to produce a DMP draft for WSDOT, specifically targeting research data. Results were presented through a talk, "State DOT Perspectives on Research Data Management," at the Transportation Research Board meeting with WSDOT sponsor Kathy Szolomayer.
Github (private) | Blog #1 | Blog #2

Transforming U.S. Census Data for the Seattle Public Library 2019
Student: Karalyn Ostler, Master of Library and Information Science Program
Sponsor: Seattle Public Library (SPL)
Developed actionable outreach intelligence, by gathering and visualizing external data resources to enhance local decision-making by staff. Conducted interviews with librarians to determine information needs, identified relevant data, cleaned data, documented findings, and

visualized socio-demographic data as a map-based dashboard using R and ShinyApps. Published a co-authored, peer reviewed paper with Bree Norlander and Dr. Weber, which was published in *Public Library Quarterly* in 2020. The work inspired new work in the Technology and Social Change group on a funded grant proposal to IMLS.
Github | Blog #1 | Blog #2 | SPL Seattle Census Data ShinyApp

Assessing Dataset Quality 2019
Student: Andrew Mckenna-Foster, Master of Library and Information Science Program
Sponsor: Washington State Library (WSL)
Worked on quality assessment for 500 datasets on data.wa.gov. Measured user perceptions of the portal with the aim of improving open data circulation. Using research literature, user interviews, and dataset analysis, researched and created a plan for removing data from civic data repositories. In May 2020, presented "The Complicated Problem of Closing Open Data" a csv,conf,v5 with WSL sponsor and former ODL intern, Kathleen Sullivan.
Github | Blog #1 | Blog #2

Improving Metadata for Seattle's Open Data 2019 (Part-time)
Student: Kaitlin Throgmorton, Master of Library and Information Science Program
Sponsor: City of Seattle
Improved metadata and data quality of City of Seattle's more than 200 open datasets. In addition to improving column descriptions (attribute descriptions / annotations), conducted outreach with data publishers and data champions to populate missing descriptions and summarize data quality issues across the portal. Updated more than 30 datasets, and reported on and proposed solutions for various data quality issues.
Github (private) | Blog #1 | Blog #2

Crowdsourcing Library Open Data 2020 (Part-time)
Student: Jamie Ramos, Master of Library and Information Science Program
Sponsor: Washington State Library (WSL)
Documented the best practices for crowdsourcing library data during an emergency, using the COVID-19 pandemic as a case study.
Github (private) | Blog #1 | Blog #2

Capstone Projects

Six teams of students worked with external partners to address open data challenges. In each respective year, their work began in January and culminated in a Capstone poster presentation at the end of May.

Opening Up the Data: Visualizing the effectiveness of Puget Sound restoration efforts 2017

Students: Tim Blankemeyer, Katrina Gertz, Emma Clarke
Sponsor: Puget Sound Partnership
Project Description: Interoperability, dispersed data, and inconsistent data formats are common issues across information science. This Open Data Literacy (ODL) Capstone tackled these issues in the realm of environmental restoration. Numerous restoration projects have been undertaken throughout the Puget Sound, but connecting investments in these projects to co-located indicators of habitat viability is challenging. The capstone team leveraged open, found data and open-source tools to build a scalable, sustainable data processing pipeline and an interactive, web-based visualization prototype. This helped their partners better tell the story of Puget Sound restoration efforts and demonstrates that open-source tools can help data curators meet open-data needs.
Blog post

Total Resource Observation Utility Team 2017
Students: Amanda Cummings, Edmund Deng, Jeffrey Seward, Vichit Sitthideth
Sponsor: Washington Department of Fish and Wildlife
Project Description: The Total Resource Observation Utility Team (TROUT), created a data visualization using Washington Department of Fish & Wildlife's (WDFW) high-resolution land change data. The interactive visualization makes this important data understandable to the public, policymakers, and environmental scientists. TROUT unlocked this public data and improved access by disseminating in a clear, concise, user-centered dashboard presenting actionable information. This visualization allows the user to see actual changes in the Puget Sound over time with data-rich GIS information describing the changes. TROUT's use of open-source technologies and code documentation mean the dashboard can be replicated for displaying future public spatial datasets.

WA State Public Meetings System 2017
Students: Jiayi Wang, Tin-Chia Lin, Yun Hsiao
Sponsor: Washington State Office of the Chief Information Officer
Project Description: The Capstone team working on the Washington State Public Meetings System set a data standard for all the public meetings digital records in Washington State. They created a cloud dataset based on this data standard and populated the dataset with test data scraped from the city and county council websites. They also built a website to visualize the database at its backend. The end result of the project is a platform for users to perform queries about public meetings via the website.

Waste 2 Resources 2017
Students: Gauri Chitre, Apoorva Deshpande, Varuna Damodaran, Meeta Pandit
Sponsor: State of Washington Department of Ecology
Project Description: The Washington State Department of Ecology works to manage and conserve public resources and collects information about the effectiveness of its conservation programs. The Waste2Resources project helped the Department of Ecology by transferring this information to the open data portal of Washington state (data.wa.gov) and designing

user-friendly data visualizations. This project created positive social impact by presenting the existing waste management information through interactive and powerful visualizations to the public. It will helped the department and government officials in tracking the progress of their conservation efforts more effectively.

Improving Trust & Interoperability: Metadata for Data Refuge's Open Data Catalog 2020
Students: Sam Buechler, Joan Hua
Sponsor: Data Refuge, Carole Palmer (UW iSchool)
Project Description: The Data Refuge Data Catalog archives federal climate and environmental data. It provides historical snapshots of datasets released on government data portals, which are vulnerable to deletion. Yet the metadata associated with these Data Refuge records have been minimal, and their relationship with the source records have not been clearly defined. To address this, the team investigated crosswalking solutions, improved metadata of target datasets, customized an extensible schema, standardized tagging with controlled vocabulary, and documented workflow for future-phase implementation. The results improved trustworthiness and interoperability, facilitated more seamless data discovery and retrieval, and met the needs of both archivists and researchers. (Capstone information on iSchool site.)

Open Data Wagon 2020
Students: Kaitlin Throgmorton, Erica Husting
Sponsor: Washington State Library
Project Description: Though public libraries have shared internal data, such as circulation data, publicly via open data portals, bookmobile data sharing has been limited. Sponsored by the Washington State Library, and using data from North Central Regional Library as a pilot, the Open Data Wagon project researched, collected, and published library bookmobile data openly on data.wa.gov, along with a reusable dataset template. Because bookmobile operations can be expensive, this project encouraged information sharing among libraries, expanded funding opportunities for mobile services by supplying additional data, and heightened the value of library mobile services. More info: https://opendatawagon.github.io (Capstone information on iSchool site.)

Targeted Student Development

Project coordination and research assistance
An Yan, a PhD student in the UW iSchool served as the project manager and research assistant of during Year 1. She also completed a teaching practicum working with Dr. Palmer as part of the initial ODL work to revise the curriculum of the Fundamentals of Data Curation course to include readings, lab work, and assignments related to open data curation.

Kaitlin Throgmorton served as the ODL's Research Assistant for two years. She also completed an internship in 2019 and an independent study in Winter Quarter 2020 with Dr. Palmer. Kaitlin

managed the Public Libraries and Open Data survey project, which included a literature review of public library and local government interaction and collaboration, as well as follow-up interviews with survey respondents and open data experts.

Master's student independent studies
Leili Slutz and Rochelle Lundry, MLIS students at the UW iSchool, completed independent studies working with Dr. Weber. Leili has helped design and implement a research project studying the open data services offered by USA public libraries (a summary of her work can be found here: https://github.com/OpenDataLiteracy/OpenDataAndPublicLibraries) Rochelle, who holds a JD specializing in Intellectual Property law, has developed a method for retrieving and analyzing license information for open data repositories used by cities in the USA. Her work contributed to a Capstone project in Spring 2018.

Karalyn Ostler, 2019 ODL intern and an MLIS student at the UW iSchool, completed an independent study in Autumn Quarter 2019 with Dr. Weber. Karalyn documented her internship work by preparing to publish a paper and disseminate code on transforming external open data into actionable intelligence for library frontline staff. Karalyn's internship project used R to transform open demographic information, such as census data, into a dashboard of interactive maps library staff could use for more targeted user outreach, and her independent study explored how to translate computational work into reproducible, publishable research using a container method. (Project Github.)

Undergraduate student experience
Jackson Brown, an undergraduate major in the iSchool's Informatics program, completed a research project with Dr. Weber in building an API to retrieve, and transform Seattle City Council data from an existing open data portal managed internally by the City of Seattle. Jackson's work is contributing to an ongoing personal research project in building civic technologies to support increased government accountability, as well as a curation workflow for improving city council data that will be used by the City of Seattle's IT department.

Selected Alumni Placements

Interns:

Lillian Curanzy: Cataloguing and Outreach Librarian, Newport Public Library (Newport, OR)

Joan Hua: Media Asset Manager, KEXP (Seattle, WA)

Andrew McKenna-Foster: Product Specialist (Biological Sciences; Information and Computing Sciences; Earth and Environmental Sciences), Figshare (Eugene, OR)

Kathleen Sullivan: Open Data Literacy Consultant, Washington State Library

Kaitlin Throgmorton: Bioinformatics Analyst/Data Curator, Sage Bionetworks (Seattle, WA)

Other representative placements:

Amy Trost: Data Services & Business Librarian, University of Maryland Libraries (College Park, MD)

Kara Woo: Senior Data Science Engineer, InsightRX (San Francisco, CA), previously Principal Bioinformatics Engineer, Sage Bionetworks (Seattle, WA)

Partner Engagement

To closely align our work with public sector perspectives and needs, ODL engaged multiple partners from state and local government as well as public libraries. These relationships were fundamental to building capacity and assuring practical applicability of our interdependent work plan of education, engagement, and research activities. Library partners represented urban, rural, and state libraries; government partners included both state agencies and city government. The government partners were essential to understanding aspects of the open data movement from the inside, where data resources are generated. They offered a direct connection with current practices and challenges from the perspective of data producers striving to meet new demands for government transparency. Library partnerships allowed us to contribute to and learn from plans, programs, and services emerging within individual libraries, assets that were greatly extended through our partnership with the Washington State Library (WSL). Formal partners:

- Washington State Library
- Washington State Office of the Chief Information Officer
- Washington State Department of Transportation
- Asotin County Library
- Seattle Public Library
- City of Seattle

From 2017-2020, we successfully engaged with an established set of partners to develop three summers of internship projects (as well as two remote internships in the spring, one in 2019 and one in 2020), described below. We also made important progress on increasing future engagement with public libraries through a new partnership with the Washington State Library (WSL) and additional interactions with our partner Will Saunders, at the Washington State Office of the Chief Information Officer, in his leadership role with the Knight Foundation funded project, Data Equity for Main Street. In particular, after a highly successful internship placement at WSL, the State Librarian, Cindy Aden, became an enthusiastic ally and partnered with Mr. Saunders to explore increasing personnel to work on open data efforts in the WSL. Through two more internships, we strengthened our relationship with WSL, as a site for placing interns but also as

a collaborator for facilitating broader outreach to public libraries across the state. WSL also facilitated an introduction to Asotin County Library (ACL) in Asotin, Washington, and Asotin sponsored one ODL intern for multiple quarters, extending the project's reach into rural public libraries and allowing the intern to create ACL's first open library dataset. Cindy Aden and other professionals at WSL were also invaluable advisors for developing educational resources with direct application to current public library practice. WSL also collaborated with us on a statewide survey of library capacity for open data initiatives, which will serve as a model for a possible national survey.

In June 2020, we hosted a virtual reunion event for partners and interns to recognize the students and sponsors contributing to and benefiting from the Open Data Literacy Project's internship program. Six sponsors, six interns, and all members of the ODL team attended. The program included a retrospective of the program, lightning talks by students on what they learned or gone on to do since their internship, reflections from sponsors on the internship process and experience, and a discussion of future directions. +

Public Libraries and Open Data Survey

In 2019, ODL surveyed public library systems and branches in Washington state, in collaboration with Washington State Library, to determine their open data activities and interests. The survey was distributed online using Qualtrics, and contained eleven questions that asked respondents about current activities related to open data, activity/interest level in open data activities (such as reference work, publishing, and building open data collections), capacity and resource challenges, open data alignment with library priorities, and support needed to move forward with open data. The initial survey invitation was sent by the Washington State Librarian to all public library systems in the state, as well as some branches. Five of the 15 largest systems agreed to distribute the survey to their branch managers. A total of 114 invitations were sent to 60 systems and 54 branches; 51 surveys were completed for a 45% response rate. The survey found that open data activity at Washington public libraries was higher than expected, particularly at reference and with libraries publishing open data on their own operations. The team has engaged with other states libraries interested in extending the survey to in their state. We have shared the instrument with one state library and the questionnaire and data will be shared in an open repository. See Appendix 4 for the survey instrument and a frequency report of results.

Preparing and Publishing Library Data: Public Curriculum

To fill a gap in open data educational resources available for practicing professionals, ODL prepared a curriculum that guides users through the process of preparing and publishing data openly. While the curriculum can be used by anyone, examples focus on library data. The curriculum reviews data, data types, formats, and structures, walks through the process of identifying data fit for publishing, offers data cleaning best practices, addresses potential privacy

issues, and finally, discusses documentation and the final act of publishing a dataset. The curriculum is publicly available on Github: (https://opendataliteracy.github.io/Prepare_Publish_Library_Data/).

Outreach and Dissemination

Note: In the Dissemination section, student authors are identified with bold text.

Curriculum Outputs

Weber, N. (2020). *Data Curation II.* https://nniiicc.github.io/LIS-598-DC2-Sp2020/

Weber, N. & Norlander, B. (2020). *Preparing and Publishing Library Data*. https://opendataliteracy.github.io/Prepare_Publish_Library_Data/

Norlander, B. (2018). How to use an API (and what is it anyway?) [Video].

Yan, A., Norlander, B., & Palmer, C. (2019). *APIs for Open Government Data.* http://odl.ischool.uw.edu/api_tutorial/

Yan, A. (2018). *Working with civic data portal metadata with OpenRefine.* http://odl.ischool.uw.edu/openrefine_tutorial/

Published Papers

[* indicates peer reviewed]

Palmer, C., Weber, N., **Throgmorton, K.**, & Norlander, B. (2021). Public Libraries and Open Government Data: Partnerships for Progress. *Library Journal*. https://www.libraryjournal.com/?detailStory=public-libraries-and-open-government-data-partnerships-for-progress

Abstract: Public libraries have long served as infomediaries between government and the public. With the acceleration of open data, libraries can expand the scope of their infomediary capacity to include providing access to valuable government information. Recognizing the opportunities and the challenges of this new area of responsibility for public libraries, the Open Data Literacy (ODL) project has been fostering the advancement of open data for the public good through education informed by engagement and research. This paper provides an overview of ODL work to date, including highlights from a survey of the current landscape of open data in public libraries in the state of Washington that shows that libraries

large and small, urban and rural, recognize the importance of open data for their communities, and that many are already integrating open data into the purview of their service commitments.

***Ostler, K. R.**, Norlander, B., & Weber, N. (2020). Using Open Data to Inform Public Library Branch Services. *Public Library Quarterly*, https://doi.org/10.1080/01616846.2020.1798206

Abstract: This article describes the curation and use of open demographic data to inform public library services. A case study of census data curated for the Seattle Public Library (SPL) system is described. To understand the information needs of library branches, a set of SPL regional managers were interviewed, a set of use cases were created, and a prototype dashboard tool using open census data was developed to address the needs of two SPL regions. The utility of available open data to meet the needs of regional managers is reviewed, as well as the potential development of replicable data analysis tools for keeping public libraries aware of shifting neighborhood demographics.

***Throgmorton, K.F.**, Norlander, B., & Palmer, C.L. (2020). Open data in public libraries: Gauging activities and supporting ambitions. *Proceedings of the Association for Information Science and Technology, 2020, 57(1)*:e321. https://doi.org/10.1002/pra2.321

Abstract: As the open data movement grows, public libraries must assess if and how to invest resources in this new service area. This paper reports on a recent survey on open data in public libraries across Washington state, conducted by the Open Data Literacy project (ODL) in collaboration with the Washington State Library. Results document interests and activity in open data across small, medium, and large libraries in relation to traditional library services and priorities. Libraries are particularly active in open data through reference services and are beginning to release their own library data to the public. While capacity and resource challenges hinder progress for some, many libraries, large and small, are making progress on new initiatives, including strategic collaborations with local government agencies. Overall, the level and range of activity suggest that Washington state public libraries of all sizes recognize the value of open data for their communities, with a groundswell of libraries moving beyond ambition to action as they develop new services through evolution and innovation.

*Weber, N., & Norlander, B. (2019). Open data publishing by public libraries. *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL), 2019*, 158-161. https://doi.org/10.1109/JCDL.2019.00031

Abstract: Public libraries in the USA are part of a broad civic information ecosystem that is rapidly adopting transparency legislation aimed at publishing structured open data for public reuse. In this preliminary results paper we look specifically at the open data publishing practices of 85 public libraries in the USA. We find that less than half of these libraries have published any open data, and that there is no relationship between revenue nor staff size and open data

publishing practices. Categorizing public library open data by type we find overwhelmingly the most frequent type of open data published by libraries are geospatial (map) information. We use these findings to develop a proposal for public libraries to engage in publishing a core set of open data, and conclude by discussing the potential for reuse of open public library data.

Throgmorton, K., Norlander, B., & Palmer, C. (2019). Open Data Literacy and the Library. *Alki: The Washington Library Association Journal, 36*(2), 27–29.

Abstract: In the digital era, *information* literacy requires *data* literacy. As we read, work, play, and explore online, we are constantly consuming data while also contributing our own data to the systems and platforms we are using. We need data skills to understand the increasingly data-driven information and decisions we encounter every day, but also to make use of the array of data resources available to us. More and more data are being released openly online for access and use by anyone. In response, libraries are becoming active in this open data movement, promoting data literacy and the use of data for the public good. Here we share perspectives from the Open Data Literacy project team on open data, why it is important, and how to get involved in open data activities in your library.

Yan, A., & Weber, N. (2018, March). Mining open government data used in scientific research. In *International Conference on Information* (pp. 303-313). Springer, Cham.

Abstract: In the following paper, we describe results from mining citations,mentions, and links to open government data (OGD) in peer-reviewed litera-ture. We inductively develop a method for categorizing how OGD are used by different research communities, and provide descriptive statistics about the publication years, publication outlets, and OGD sources. Our results demon-strate that, 1. The use of OGD in research is steadily increasing from 2009 to2016; 2. Researchers use OGD from 96 different open government data por-tals, with data.gov.uk and data.gov being the most frequent sources; and, 3.Contrary to previous findings, we provide evidence suggesting that OGD from developing nations, notably India and Kenya, are being frequently used to fuel scientific discoveries. The findings of this paper contribute to ongoing research agendas aimed at tracking the impact of open government data initiatives, and provides an initial description of how open government data are valuable to di-verse scientific research communities.

Presentations

*Acker, A., Donaldson, D., Kriesberg, A., Thomer, A. and Weber, N. (2020). Integrating Research and Teaching for Data Curation in iSchools. *Proceedings of the Association for Information Science and Technology, 2020, 57(1)*:e285. https://doi.org/10.1002/pra2.285

Abstract: The quickly changing nature of information science and technology creates unique and remarkable challenges in terms of developing curriculum focused on building data competencies. Faculties responsible for teaching current developments in information studies

have the unique burden of needing to continuously update our curricula without sacrificing our broader teaching goals. This panel features diverse perspectives on teaching data curation skills in five US-based schools of information at the undergraduate and graduate levels. Panelists will present their unique perspectives on pedagogical approaches in courses dedicated to data curation, digital preservation, description and access standards, as well as data access and interchange. Topics introduced will range from flipped classroom techniques, finding messy datasets, common pitfalls, hands-on labs, cloud based tools, data carpentry labs, and sequencing learning objectives to match stages of the data life cycle. This panel will give ASIST conference participants an opportunity to see a range of junior faculty, each with IMLS funded research projects related to data curation, share their experiences of teaching data competencies in the classroom.

Sullivan, Kathleen, and **Mckenna-Foster, Andrew.** (2020, May 13). The Complicated Problem of Closing Open Data. Presented at csv,conf,v5. http://doi.org/10.5281/zenodo.3823409

Hua, Joan. (2020, January). Transportation Research Data Management at Washington State DOT. Transportation Research Board Meeting. Co-presentation with Kathy Szolomayer. (Presentation information.)

Palmer, C.L., Norlander, B., & **Throgmorton, K.F.** (2019, October 1). *What data does your community need the most?* Webinar presented at Washington State Library's First Tuesdays, Online. https://my.nicheacademy.com/washingtonstate/course/9291

Abstract: As government publishes more of its data online, libraries can provide expert guidance about what information to prioritize. But what data does the community need most, and what are the challenges to getting that data published? Hear what public libraries in Washington State are doing to get community-centered data online, and learn about the latest research on local data needs.

Palmer, C.L., and Aden, C. (2019, November 4). From Ambition to Action: State and Public Libraries Partnering on Open Data. Chief Officers of State Library Agencies (COSLA) Fall Meeting, Hartford CT.

Hua, J. (2019, October). Transportation Research Data Management: A Collaboration through the Open Data Literacy Project. Digital Library Federation Forum. Lightning talk.

*Palmer, C. L., Weber, N., Greenberg, J., Lin, X., Howison, J., Oh, S., & Murillo, A. (2018). *Confronting the Expanse of Data Education: From Local Open Data to Global Cyberinfrastructure*. Panel presentation. Annual Conference of the Association for Information Science & Technology, Vancouver, BC.

Abstract: As the demand for data science and data-intensive capabilities grows in all sectors, educators in schools of information and library and information science are working to

deepen and expand their programs to meet workforce expectations. This panel will examine current trends and investments in data education and professionalization, with an emphasis on the unique contributions information professionals bring to the data workforce. Educators will present on current initiatives that represent the breadth and complexity of preparing information professionals for work in data intensive environments. Their unique approaches anticipate and respond to workforce demand, address significant educational challenges, and offer models for making progress within the varying contexts of schools in different regions of the U.S. and Asia. Half of the session will be reserved for audience engagement designed to leverage and share the wealth of experience of the educators and students in attendance. The exchange will generate ideas about new directions and successful approaches that educators can apply to set priorities, position their programs, and collaborate to enrich and provide leadership in data education.

Note: Dr. Palmer served as the panel organizer and facilitator and provided a framing introduction to the session, highlighting new directions, opportunities, and responsibilities for educating information professionals and the range of approaches and innovations represented by the group. Dr. Weber addressed outcomes from the first two years of ODL student experiences working with open data, and the implications of these experiences for curriculum development in the graduate programs at the UW iSchool, citation below.

Weber, N. 2018. *Preparing Masters Students for Public Sector Data Service*. Annual Conference of the Association for Information Science & Technology, Vancouver, BC.

Palmer, C. L., Weber, N. & Johnson, T. (2017, April 14). *An Educational Workbench for Data Curation*. Workshop presented at 7th National Data Service Consortium Workshop, Chicago, IL.

Abstract: The Data Curation Educational Workbench provides a platform for students to gain hands-on experience with data curation software and tools. The workbench is being piloted with students at the University of Washington (UW) Information School enrolled in the Fundamentals of Data Curation and Advanced Data Curation course. We will provide guidance for students to develop, publish, and manage collections of datasets through the platform's user interfaces and APIs. Students will use the workbench to experiment with and apply tools to a range of curation tasks. The initial phase will focus on building a coherent set of tools around stable instances of CKAN and Dataverse for building collections and complementary tools for performing curation tasks, such as FIDO, SQL Workbench, Oxygen, EIS Archiver, and selected harvesting tools.

Other presentations:

Carole Palmer presented an Open Data Literacy Retrospective at the ODL Reunion, June 5, 2020.

Kaitlin Throgmorton presented two short overviews of the Open Data Literacy project, first to the UW Provost on April 15, 2019, and to the All iSchool Welcome on September 19, 2019.

Bree Norlander presented a short overview of the Open Data Literacy project on March 13th, 2018 at the Monthly OpenCon Librarian Community Call[1]. Notes from the call are available at: http://bit.ly/ECLCall.

Posters

Palmer, C., Weber, N., Norlander, B. & **Throgmorton, K.** (2020) *Expanding Open Data Expertise in the Public Sector*. University of Washington iSchool Research Fair. Canceled due to pandemic.

Abstract: Open Data Literacy (ODL) improves accessibility and use of open data by advancing data literacy and expanding open data expertise in the public sector. To do this, ODL has developed iSchool graduate curriculum, coordinated open data fieldwork, produced open educational resources, analyzed library data published on city open data portals, and surveyed public libraries across the state on current activities, ambitions, and challenges in open data.

*Weber, N., Palmer, C. & Norlander, B. (2018) *Advancing Open Data: Aligning Education with Public Sector Data Problems*. Annual Conference of the Association for Information Science & Technology, Vancouver, BC.

Abstract: The Open Data Literacy (ODL) project is working to prepare students with the open data competencies needed in public libraries and government agencies, by advancing curriculum and providing field experiences in collaboration with Washington state agencies. To determine curriculum gaps, we are examining data problems faced by public sector organizations. Based on the first year of student field experiences, we found that data organization and access, data quality, and data visualization were prominent problem areas. While captured by current frameworks of data expertise, certain competencies, such as data organization and access, need to be prioritized and deepened in the curriculum.

[1] https://www.opencon2018.org/community_calls

*Weber, N., **Yan, A.**, & Palmer, C. L. (2017, March 23). *Open Data Literacy Project*. iConference 2017. Wuhan, China. https://doi.org/10.9776/17341

Abstract: The Open Data Literacy project is preparing future and current librarians to advance open data initiatives in public libraries. This poster will provide an overview of our core activities and progress to date, with a focus on strategies iSchools can implement to collaborate with public sector partners to overcome the current lag in data expertise in the public library workforce. Core activities include new curriculum for master's students in Library and Information Science, a slate of fieldwork opportunities at institutions with open data initiatives, and community workshops and open education resources for public librarians and information professionals. The educational framework will improve public accessibility and use of open data while increasing the data capabilities of both new and practicing information professionals in public libraries.

Weber, N., **Yan, A.**, & Palmer, C. L. (2017, February 21). *Open Data Literacy Project*. University of Washington eScience Institute Poster and Networking Session. University of Washington, Seattle.

*Weber, N. & **Yan, A.** (2017) *Integrating User Feedback with Open Data Quality Models*. Annual Conference of the Association for Information Science & Technology, Washington, DC.

Abstract: User feedback is critical to improving the quality of open data. However, most open data publishers gather only anecdotal evidence about user experience. This unstructured and informal commentary is, consequently, difficult to translate into actionable steps towards improving data quality. Drawing on user comments collected from Data.Gov - an open data portal providing access to thousands of datasets published by city, state, and federal government agencies in the USA – we inductively develop a classification of reported data quality issues. This poster presents preliminary findings from applying this coding scheme to all issues that users filed on Data.gov in 2015 and 2016. We suggest that our classification scheme can help open data publishers collect structured, actionable information to improve data quality.

Medium Blog

An ODL publication on Medium.com (https://medium.com/open-data-literacy/archive) was set up for interns to publish articles and build awareness of Open Data Literacy projects. The ODL team also disseminated information about ongoing research and open data news.

Posts by ODL interns:

Ramos, Jamie. (2020, June 23). COVID Data Collection: Lessons Learned and Future Approaches. *Medium*. https://medium.com/open-data-literacy/covid-data-collection-lessons-learned-and-future-approaches-8bfc3420ec0d

Ramos, Jamie. (2020, April 30). Tracking Washington State Public Library Responses in Crisis. *Medium.* *https://medium.com/open-data-literacy/tracking-washington-state-public-library-responses-in-crisis-60482ea7e080*

Ostler, Karalyn. (2019, September 20). Using Open Data to Understand Communities Around Seattle Public Libraries. *Medium*. https://medium.com/open-data-literacy/using-open-data-to-understand-communities-around-seattle-public-libraries-fa3f9a96f3be

Hua, Joan. (2019, September 14). Laying the Groundwork for Transportation Research Data Management. *Medium*. https://medium.com/open-data-literacy/laying-the-groundwork-for-transportation-research-data-management-b8fd222a9d52

Curanzy, Lilian. (2019, September 4). Opening Rural Library Data: Applying Coursework to the Field. *Medium*. https://medium.com/open-data-literacy/opening-rural-library-data-applying-coursework-to-the-field-3137ce64c187

Mckenna-Foster, Andrew. (2019, August 25). Facing the Flood: Assessing Metadata Quality on Washington's Open Data Portal. *Medium*. https://medium.com/open-data-literacy/facing-the-flood-assessing-metadata-quality-on-washingtons-open-data-portal-d76312d767fe

Throgmorton, Kaitlin. (2019, August 9). The Myth of the Obvious. *Medium.* https://medium.com/open-data-literacy/the-myth-of-the-obvious-ca6afb08b02d

Ostler, Karalyn. (2019, July 24). Beyond the Census: Using Census Data in Public Libraries. *Medium.* *https://medium.com/open-data-literacy/beyond-the-census-using-census-data-in-public-libraries-333e2643fd21*

Mckenna-Foster, Andrew. (2019, July 19). The Complicated Problem of Dataset Removal. *Medium*. https://medium.com/open-data-literacy/the-complicated-problem-of-dataset-removal-e79e97a23b48

Curanzy, Lilian. (2019, July 9). What's Not There: A case for public missing item reports. *Medium*. https://medium.com/open-data-literacy/whats-not-there-c058a933e1bf

Gonzalez-Curci, Angela. (2018, August 9). WSDOT Records as Open Data. *Medium*. https://medium.com/open-data-literacy/wsdot-records-as-open-data-731d525e0af3

McCraney, Kevin. (2018, August 9). Summarizing a Summer Devoted to Data. *Medium*. https://medium.com/open-data-literacy/summarizing-a-summer-devoted-to-data-6cf7f68f29a1

Denning, Leslie. (2018, August 8). Wrapping up with Public Records & the Open Data Alliance in Washington. *Medium*. https://medium.com/open-data-literacy/wrapping-up-with-public-records-the-open-data-alliance-in-washington-493b1ac93c24

Sullivan, Kathleen. (2018, August 8). Can Libraries Accelerate Local Open Data Publishing? *Medium*. https://medium.com/open-data-literacy/can-libraries-accelerate-local-open-data-publishing-34caaf6a767f

Gonzalez-Curci. (2018, July 12). A Stake in the Game: Public Disclosure Coordinators at the Open Data Table. *Medium*. https://medium.com/open-data-literacy/a-stake-in-the-game-public-disclosure-coordinators-at-the-open-data-table-bdcc1fa5c7d7

Sullivan, Kathleen. (2018, July 11). Need help with open data? Ask a librarian. *Medium*. https://medium.com/open-data-literacy/need-help-with-open-data-ask-a-librarian-bb55b5477334

McCraney, Kevin. (2018, July 10). Improving User Experience for Public Records Requests. *Medium*. https://medium.com/open-data-literacy/improving-user-experience-for-public-records-requests-be311e302b7a

Denning, Leslie. (2018, July 10). The Public Records Request Life Cycle. *Medium*. https://medium.com/open-data-literacy/the-public-records-request-life-cycle-cacb5be692f

Tranfield, M. Wynn. (2017, September 24). Data Equity Curriculum Building. *Medium*. https://medium.com/open-data-literacy/data-equity-curriculum-building-867a2e4539cc

Blankenmeyer, Tim. (2017, September 22). ODL project builds foundations for future of innovative WSDOT data catalog. *Medium*. https://medium.com/open-data-literacy/odl-project-builds-foundations-for-future-of-innovative-wsdot-data-catalog-64133bb1906f

Showell, Nina P. (2017, August 19). Using Metadata Standards at the City of Seattle. *Medium*. https://medium.com/open-data-literacy/using-metadata-standards-at-the-city-of-seattle-4ce93869a2

Showell, Nina P. (2017, July 24). Metadata Standards and Data Usability at the City of Seattle. *Medium*. https://medium.com/open-data-literacy/metadata-standards-and-data-usability-at-the-city-of-seattle-773451af4ee9

Carrier, Sarah. (2017, July 18). Seattle Public Library is 'Up and Running.' *Medium*. https://medium.com/open-data-literacy/spl-is-up-and-running-927cf79eed38

Blankenmeyer, Tim. (2017, June 30). ODL project helping WSDOT bridge gap between data and knowledge. *Medium*. https://medium.com/open-data-literacy/odl-project-helping-wsdot-bridge-gap-between-data-and-knowledge-2dc1c03f0ca4

Carrier, Sarah. (2017, June 30). Engaging with Open Data: Seattle Public Library's Role in Open Data. *Medium*. https://medium.com/open-data-literacy/engaging-with-open-data-seattle-public-librarys-role-in-open-data-a8625859998e

Showell, Nina P. (2017, June 27). Context for Communities: Seattle and Metadata Standards for Civic Data. *Medium*. https://medium.com/open-data-literacy/context-for-communities-seattle-and-metadata-standards-for-civic-data-9e631291b146

Blankenmeyer, Tim. (2017, June 26). Open source + open data helps Capstone project visualize Puget Sound restoration. *Medium*. https://medium.com/open-data-literacy/open-source-open-data-helps-capstone-project-visualize-puget-sound-restoration-149ea12d7657

Posts by ODL team:

Norlander, Bree. (2020, May 20). 10 Recommendations for Creating Internships with the Public Sector. *Medium.* *https://medium.com/open-data-literacy/10-recommendations-for-creating-internships-with-the-public-sector-518769ba5e55*

Throgmorton, Kaitlin. (2020, April 7). COVID-19 Open Data Round-up. *Medium*. https://medium.com/open-data-literacy/covid-19-open-data-round-up-71936b75e2bd

Throgmorton, Kaitlin. (2020, March 9). Public LIbraries & Open Data Survey: Initial Findings. *Medium*. https://medium.com/open-data-literacy/public-libraries-open-data-survey-initial-findings-f5d0e666f5aa

Throgmorton, Kaitlin. (2020, October 11). What makes open data good, and how can we make it better? *Medium.* *https://medium.com/open-data-literacy/what-makes-open-data-good-and-how-can-we-make-it-better-c37275e71319*

Throgmorton, Kaitlin. (2019, July 11). Meet ODL's 2019 Interns. *Medium*. https://medium.com/open-data-literacy/meet-odls-2019-interns-86c17c75072e

Weber, Nic. (2018, July 13). 2018 Open Data Literacy Internships. *Medium*. https://medium.com/open-data-literacy/2018-open-data-literacy-internships-37cb92eaa142

Weber, Nic. (2017, June 20). Introducing the 'Open Data Literacy' project at the UW iSchool. *Medium.* *https://medium.com/open-data-literacy/introducing-the-open-data-literacy-project-at-the-uw-ischool-7a2555b3cf94*

Github

Many of the collaborative activities undertaken throughout the project required access to data, development of software, and central coordination of documentation. ODL's organizational Github account — https://github.com/OpenDataLiteracy — was used for this purpose, and grew to include 46 repositories. In addition, most interns documented their work using Github. Open-access repositories of note include:

- All materials and course deliverables from the INFX 551 course:
  - Spring 2017
  - Winter 2018
- Research related repositories for ongoing work with students in the iSchool
- Internship repositories that host outputs from the ODPG summer internships
- Outreach repositories for past and planned events

Website

The ODL website showcased the work of the Open Data Literacy project. It included pages for each internship and capstone project.

Information Sessions

Open Data Literacy sessions held for graduate students in November of 2016 and 2017 to introduce the Information School community to the area of professional practce and developing curriculum. Topics included:

1. Overview, Motivation, Partners
2. Curriculum (courses you can take)
3. Student Project Opportunities
4. Sample Projects
5. Opportunities at Seattle Public library

Session attracted a range of students and supported recruitment of students for Capstone Projects. In 2017, partners from Seattle Public Library, Socrata, and National Telecommunications and Information Administration were on hand to discuss potential upcoming projects with interested students.

Evaluation

Formative evaluations were conducted periodically throughout the course of the project to inform curriculum development and student field experiences, especially the internship program. Additionally, a formal summative evaluation was conducted at the close of the grant period to assess how ODL curriculum and educational experiences assisted iSchool graduates in their careers.

Summative Evaluation

The formal summative evaluation was conducted by Stacey Wedlake, a researcher from the Technology and Social Change group in the Information School. As someone knowledgeable about ODL activities but external from the overall initiative, she was well positioned to perform the evaluation work. Her full report (Appendix 1) extends the formative evaluations conducted throughout the project by providing a more complete and holistic account of student experiences and outcomes, based on a survey she distributed to 2017-2020 alumni who completed ODL coursework and/or an internship. The results provide an assessment of the value of the curriculum and field work experiences, current job placements, and how graduates are applying data expertise in their work.

The survey was distributed via email to 94 alumni in February, 2021 to students who had completed at least one of the two Data Curation courses and/or an ODL internship. It consisted of mostly closed, multiple-choice questions with open-ended response options and a final open question. Results provide a longer-term assessment of the ODL value of the curriculum and field work experiences, current job placements, and how graduates are applying data expertise in their work .

The response rate was 23%. In our analysis, we also used the iSchool Career Services Alumni Placement Survey to compare entry-level career data for alumni in the ODL sub-set.

Half of all survey respondents graduated in 2020, and ten were ODL interns. All but one of the ODL alumni were working within the year after graduation, and 70% were “fully employed” (not looking for other work). Five have received a promotion or changed jobs since they graduated and none were unemployed. Sixty-four percent reported working in either libraries or education (32% in each respective category), which are also the top two markets for all iSchool’s MLIS alumni, however ODL alumni work in a greater variety of industries.

Most ODL alumni spend a majority of their work time in a variety of activities. Data management and metadata and documentation were the highest rated activities, followed by user instruction and training, and data discovery and retrieval. In rating curriculum that contributed to their professional progress, data and metadata standards (79%) was selected most frequently, along with data management best practices (79%), and followed by data management planning (74%). They also reported strong capability for continued learning, with 78% reporting that the curriculum prepared them “extremely” or “very” well to learn new data skills.

Formative Evaluation

Course Evaluations

Each course at the iSchool concludes with a student survey administered by the Academic Dean’s office. We were unable to alter these evaluations to match IMLS Performance Measure Statements, but the evaluations still allowed us to track student experience and satisfaction with course material, level of rigor, and instructor engagement. The survey results are also utilized by the iSchool administration to evaluate course offerings. The standard evaluation survey primarily uses a 5-point rating scale, and includes four open-ended questions.

Out of 20 courses taught between 2017 and 2021, we have access to a subset of 5 course evaluations, 4 for Data Curation I and 1 for Data Curation II, taught by Carole Palmer between 2017-2020. The courses represent both in-person and online instruction. Response rates for each evaluation ranged between 50%-90%, with the lowest occurring in Spring 2020, due to the Covid-19 pandemic.

In the subset of 5 course evaluations, students reported an above-average level of engagement, and that the course met their expectations for level of difficulty. The courses received an average rating of 4.6/5 for overall quality based on a 5-point rating scale.

In the open-ended responses, students described the class as intellectually stimulating, well-organized, and as an approachable foundation of the data curation discipline. Many were particularly appreciative of the field expertise Dr. Palmer brought to the classroom. In earlier years of the survey, many students suggested a stronger emphasis on practical elements such as lab exercises, which were implemented in later course revisions.

The entry-level course, Data Curation I, frequently filled at or above capacity immediately after registration opened. As noted by the MLIS program advisor, courses that fill that quickly are seen by students as critical to one or more career paths. The advisor also noted that students also began indicating interest in data curation and digital preservation in their admissions applications. Students have consistently created impressive portfolios of data curation work through the sequence of courses. Many also participated in the ODL internship program, discussed below, or sought out data curation directed fieldwork opportunities, including capstone projects and independent study experiences.

Internship Evaluations

Student Focus Groups

Each summer intern cohort (2017, 2018, and 2019) participated in focus groups reflecting on their experiences. Each focus group took place at the end of the 8-week internship term, and was facilitated by two members of the ODL staff. Questions and probes were written collaboratively by the ODL team, and modified in subsequent years based on review of standout themes. Students discussed project deliverables and anticipated takeaways, and provided feedback on the coursework and skills that prepared them for their projects.

Nearly every intern completed at least one ODL course, and the majority completed one or both courses in the Data Curation sequence. Students in each cohort spoke to the value of the ODL curriculum for planning and execution of their projects, citing particular concepts and tools such as APIs, data management plans, and OpenRefine. They similarly noted the courses' timely content as a key factor to both their project design, and to advocating for their decisions and recommendations to their sponsors.

Students were overall satisfied with their respective placements, and felt their projects complemented their occupational backgrounds while appealing to their growing interests as information professionals. Multiple students in each cohort recounted the support of on-site mentors, who integrated students with other employees and helped familiarize them with organizational operations. Students also reflected on the opportunities for professional development facilitated by their mentors, including networking with peer agencies. One student described how participating in statewide open data committee meetings provided them a broader perspective on current open data initiatives.

Other students similarly shared how their on-site experiences shaped their perceptions of the value of open data advocacy in the public sector. Across different sponsor sites, students encountered institutional barriers that hindered adoption of open data practices, such as siloed departments and data literacy gaps within the staff. Members of each cohort described how mediating these challenges affirmed the necessary role of information professionals in public sector work, and further influenced their career trajectory.

Students overwhelmingly agreed that the internship program helped them synthesize technical skills acquired in ODL coursework, alongside other information science skills acquired in the broader iSchool curriculum. They described how direct application of class concepts in real-world environments helped both solidify the practical importance of those skills, and bolster their confidence in applying them. Students also appreciated the authentic impact of their internship projects, in contrast to theoretical course assignments. One student reflected on a feeling of professional responsibility inspired by the combination of ODL coursework and the internship program:

> "I think there's a real need for people like us to be kind of leaders in the field. People who are, one, really passionate about open data, but two, also have a great conceptual foundation and ethics and tools that you need and the skills that you need. Because one thing that I was really surprised by at [internship site] was that there didn't seem to be a kind of a sense of ownership about it. Like that, you know, this is something that we really care about and we really want to see grow and advance. So I think having more people like that would be really crucial for open data to go to the next level. So I'm excited that a lot of us really care about it. And I think there will be more of those people going forward."

Student Survey
From 2019-2020, intern alums were also asked to complete questionnaires about how the ODL internship applied to their current work. Of the interns who responded, five were involved in data-related work (titles such as open data consultant, data curator, metadata specialist, data engineer) and considered their ODL internship to have been relevant to securing and succeeding in their current work.

Sponsor / mentor Interviews

At the close of the internship program, we conducted semi-structured interviews with five sponsor organizations to assess their experiences with the ODL program. Experiences were highly positive across the sponsor group. [refer to narrative report?]

In interviews, sponsors spoke highly of the competencies, professionalism, and enthusiasm students brought to the internships. They described students' critical role in advancing organizational initiatives, highlighting students' proficiency in technical skills lacking in the workplace, and their singular focus on projects, which expedited programs that would otherwise have slowly progressed due to limited staff capacity. As one sponsor explained:

> "I was able to accomplish a couple of important goals for the state by virtue of enlisting the innovation capacity and the persuasive capacity of ODL students... it was a kind of a wonderful marriage or mashup of student innovation and agency needs. So I was able to see some agency projects get accomplished that would not have been accomplished otherwise. And students were able to bring new and innovative perspectives to the work

> that I really felt were needed. So I achieved my two objectives of number one: getting some open data work done that I couldn't do myself and that I knew agencies were challenged to actually plan or accomplish. And number two: to get you know more contemporary, up-to-date tools and methods implemented."

Another sponsor recounted an intern's contributions to their organization's open data initiative, emphasizing the student's ability to bridge new practices into the organization to help ensure the longevity of the project:

> "Compared to the rest of the project, [student] has so much expertise...we'd started the conversations but [student] went in and was able to dig down...a lot further with people. We were getting what seemed to be cooperation till we got to really now we need the data and then we kind of got a wall. And [student] was able to penetrate that better than we were. And then [student] just taught us so much about the actual creating of the data sets, cleaning them up and that kind of thing….Besides, getting those data sets up there, [student] created a tool for us to use in the future. And so that is sustainability.

While describing the immediate value of interns' individual projects, sponsors often framed the students' impact within the long-term goals and professional reputation of the organization. Three notable examples include: (1) One sponsor explained how one intern's efforts in collecting disparate civic datasets positioned the organization to move forward in a grant-funded project for a county-wide open data portal. (2) Another sponsor described how the template for research data management plans developed by one intern prepared their library department to meet imminent federal requirements, and helped them gain national recognition for their proactive initiative. (3) Another sponsor cited how two interns' comprehensive curation work for a statewide open data platform similarly helped augment the organization's national reputation as a leader in open civic data.

Many sponsors spoke to the mutual benefits of the internship program, and all of them expressed the desire for continued collaboration with iSchool students. They reflected on the positive influence of students' fresh perspectives for their organizations, and the formative value of field experience for the students as emerging professionals. Sponsors independently agreed that sustained partnership with the iSchool will continue to benefit their organizations, as well as future iSchool interns.

We separately conducted review meetings with project partners from sponsor organizations at the completion of the summer internships to gather feedback on the internship process and deliverables. In 2019 this was coordinated with an informal business lunch held before the final intern showcases in 2019.

Capstone Evaluations

Surveys and Interviews
Upon completion of the Capstone projects in 2017, we surveyed both sponsors and students about their experiences. Interviews with students were conducted in 2018.

Students:
Six students responded to the Capstone survey. Each of the fours teams was represented in the survey responses. All respondents were satisfied or very satisfied with their Capstone experience. The students benefited from and appreciated the supportive relationship with the sponsor. Respondents included comments such as “Our sponsor was very enthusiastic about our project and genuinely cared about us and our work” and “It was a great, smooth experience. Our sponsor was very helpful in providing the information that we needed”. Students faced challenges with large datasets, messy data, and interoperability between web applications.

To gain further insight into the value of data curriculum and sponsor interests in data projects, our project coordinator, Bree Norlander, interviewed several student capstone teams that used open data for their project deliverables. Nic Weber met with two of the Capstone Sponsors to get feedback on their experiences supervising project teams. All of the teams interviewed were Master of Science in Information Management graduate students or Informatics undergraduate students. The projects were highly technical in nature, requiring coding and analytics skills. Student respondents identified the 3-course Data Science course sequence and the Interactive Information Visualization course as highly important for succeeding with their projects, with a need for more coursework experience in web design or user experience.

Sponsors:
Three sponsors responded to the survey from three (out of four) different teams. All respondents were satisfied or very satisfied with their experiences and all three were willing to sponsor another Capstone. One respondent even noted in the comments field, “Great experience from start to finish. I need a higher category than ‘very satisfied.’” Sponsors advised the ODL team that Capstone teams should begin with face-to-face meetings with sponsors and that the Capstone teams would benefit from closer faculty connections.

Lessons Learned

Curriculum:
Early interviews and focus groups with students and partners highlighted the need for more practical technical content. We responded with more lab-based learning and prototype development and will continue to expand and improve this component of our instruction,

especially for the online versions of these courses. We expect that further continuing education materials should also emphasize lab based and hands-on learning.
There is also a need for further attention to the community engagement dimension of open data competencies and advocacy approaches for work on data release from public sector organizations and government agencies. COVID-19 brought this particular aspect of the ODL project into focus. We have sought to incorporate a number of different examples from the pandemic in training materials. One of the substantive contributions of our internship program was a collaboration with the WA State Library where students mapped the accessibility and interoperability of different WA libraries to a common framework so that aggregate, state level data could be efficiently and effectively shared.
We also further experimented with how to effectively build a sequence of data curation and preservation courses to help students move from concepts to action across 20 weeks of combined instruction. In piloting the Open Textbook format for the 2020 Data Curation, we have also identified areas for improvements to serve both online and residential students through a "flipped classroom" model where students read and watch lectures before engaging in face-to-face discussion and lab-based exercises with instructors. We are currently working to translate this open access content to support continuing professional education with open data (our survey results show that webinars and workshops would be the most valuable formats for delivery of open data curriculum to libraries).

Internships:
With the years of improving the internship experience, for students, sponsors, and our own project management, we developed significant understanding of how the internship format can be adopted and extended in the future, which we have documented in the guide: 10 Recommendations for Creating Internships with the Public Sector.
Feedback from students and sponsors confirmed that internships projects, when designed well, can leverage and reinforce classroom learning. For example, competencies with metadata, APIs, and data ethics were successfully applied in practice. The partners also observed that online students were uniquely equipped to work seamlessly both in-person and remotely.
The high value of the internships made evident the need to scale the benefits of hands-on instruction to give more students exposure to the organizational contexts of public sector work and experience translating classroom education into work for the public good. We demonstrated how investments in practical demos and lab exercises, as well as classroom interaction with practicing data professionals, can benefit both residential and online students.

Partners:
Our partners in public libraries and government agencies were key to the success of ODL. The relationships were fundamental to building capacity and assuring practical applicability of our interdependent work plan of education, engagement, and research activities. We found the government partners vital for understanding aspects of the open data movement from the inside, where data resources are generated. They provided an invaluable connection with current practices and challenges from the perspective of data producers striving to meet new

demands for government transparency. Library partnerships made it possible for ODL to contribute to and learn from plans, programs, and services emerging within individual libraries.

Supporting open data work in libraries:
Through our work with partner libraries and the survey of public libraries we learned that, at present, libraries in Washington state are particularly active with open data through reference services. They are also beginning to release their own library data to the public. Libraries, large and small, are making progress on new initiatives, including strategic collaborations with local government agencies. Overall, the level and range of activity suggests that public libraries of all sizes recognize the value of open data for their communities. As we've seen working with ODL partners, libraries of moderate size and resources can be highly successful in catalyzing change and innovation. While libraries identified funding as a major challenge in moving forward with new or additional open data activities, other types of resources and relationships are also a high priority. Notably, survey results showed that "new or improved collaboration with city, state, or other partners," was almost equal to the need for financial resources, followed by professional development.

Appendix 1:

Open Data Literacy Summative Evaluation

Prepared by Stacey Wedlake, Research Coordinator and Analyst
Technology and Social Change Group, University of Washington iSchool
April 20, 2021

Overview

This evaluation was designed to better understand how ODL curriculum and educational experiences assisted students in their career aims, including obtaining a professional position, making progress in their career, and supporting their day-to-day work and responsibilities. To do this, we created a survey to send out to all iSchool 2017-2020 alumni that participated in at least one of the following ODL educational experiences: Data Curation I, Data Curation II, and/or ODL internship. The survey contained mostly closed, multiple-choice questions, but did have optional open-ended responses. Questions asked alumni about their careers since graduating from the iSchool, how they apply data expertise in their work, and feedback on curriculum. The results help us understand the impact of the ODL project and how future curriculum revisions and educational experiences can best support students' careers.

We distributed the survey via email to a total of 94 alumni on February 10, 2021. Alumni names and contact information came from class rosters and ODL project communications. iSchool Alumni Relations cross-checked the contact information with their own records and sent the survey. We also reached out directly to ODL intern alumni from an ODL faculty email address. The survey closed on February 23rd, 2020, receiving 22 responses.

In our analysis, we also used the iSchool Career Services Alumni Placement Survey to compare entry career data for alumni in the ODL sub-set. This survey is open annually from May - December and sent to all alumni that graduate within that academic year. The questions ask about job search, current employment, industry, and salary. We used survey data from 2017-2020 to compare entry-level career data of all MLIS graduates.

Respondent Profile

Half of all survey respondents graduated in 2020, and 19 respondents were MLIS alumni (10 online and 9 residential). The most common educational experience of respondents was the class Data Curation I (see Figure 1). A majority of respondents participated in more than one

experience, and 1 alumni engaged in all 5 experiences. In contrast, 6 respondents only took Data Curation I.

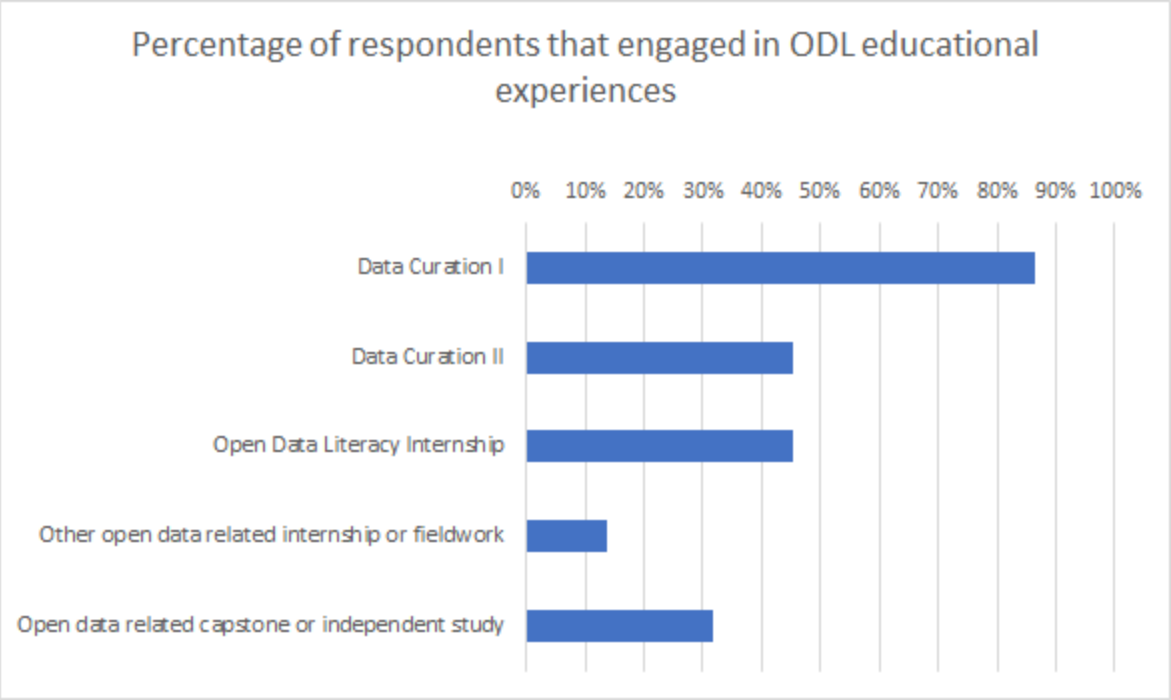


Figure 1.

Almost half of respondents (10) were ODL interns which means that 71% of all ODL alumni interns took the survey. ODL interns participated in a wider variety of educational activities (Figure 2): 6 out of 10 took both Data Curation courses and another 5 participated in an open data related capstone or independent study. One intern only participated in the ODL internship and not other activities. Two interns did not take Data Curation I but placed directly into Data Curation II.

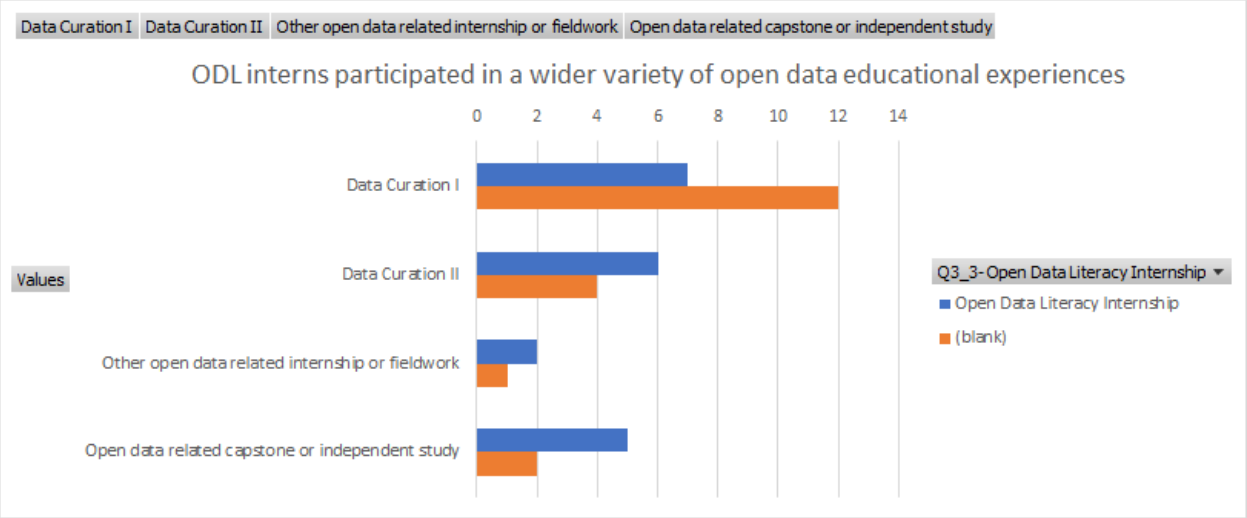


Figure 2.

ODL alumni careers

All but one ODL alumni was working within the year after graduation (Figure 3), and 70% of all alumni were “fully employed” (not looking for other work).

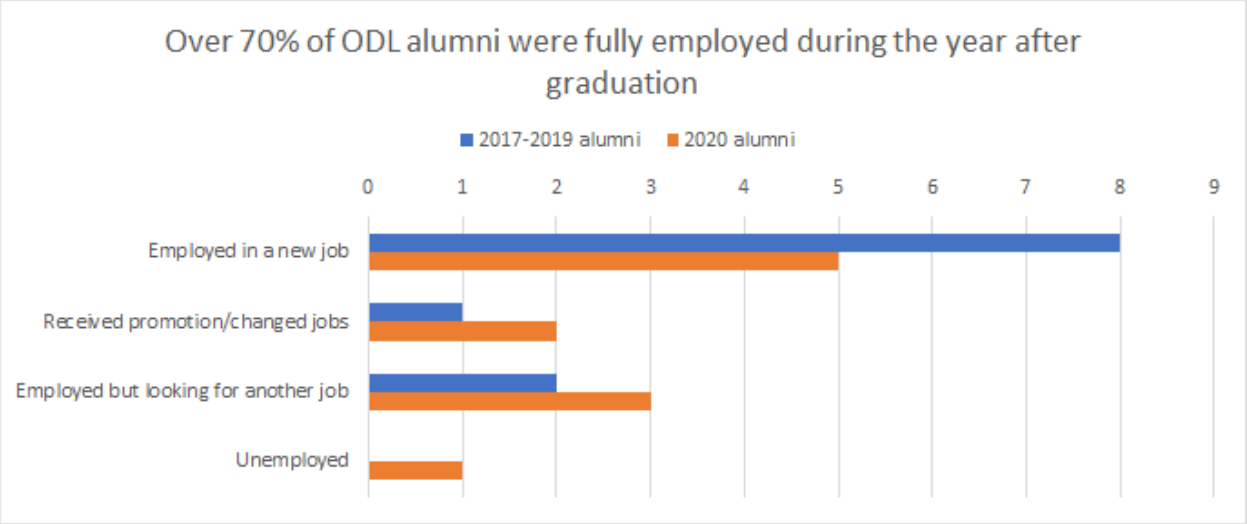


Figure 3.

We also asked 2017-2019 alumni how their employment has or has not changed since the year after graduation. Five have since received a promotion or changed jobs and none were unemployed.

To better understand the kind of work alumni ended up doing, we asked respondents to classify their industry in one or two categories. The greatest proportion of ODL alumni reported working in libraries and education (see Figure 4). These industries were also the top two industries for MLIS Alumni in the iSchool Career Services survey. However, a much higher proportion of all MLIS alumni work in libraries.

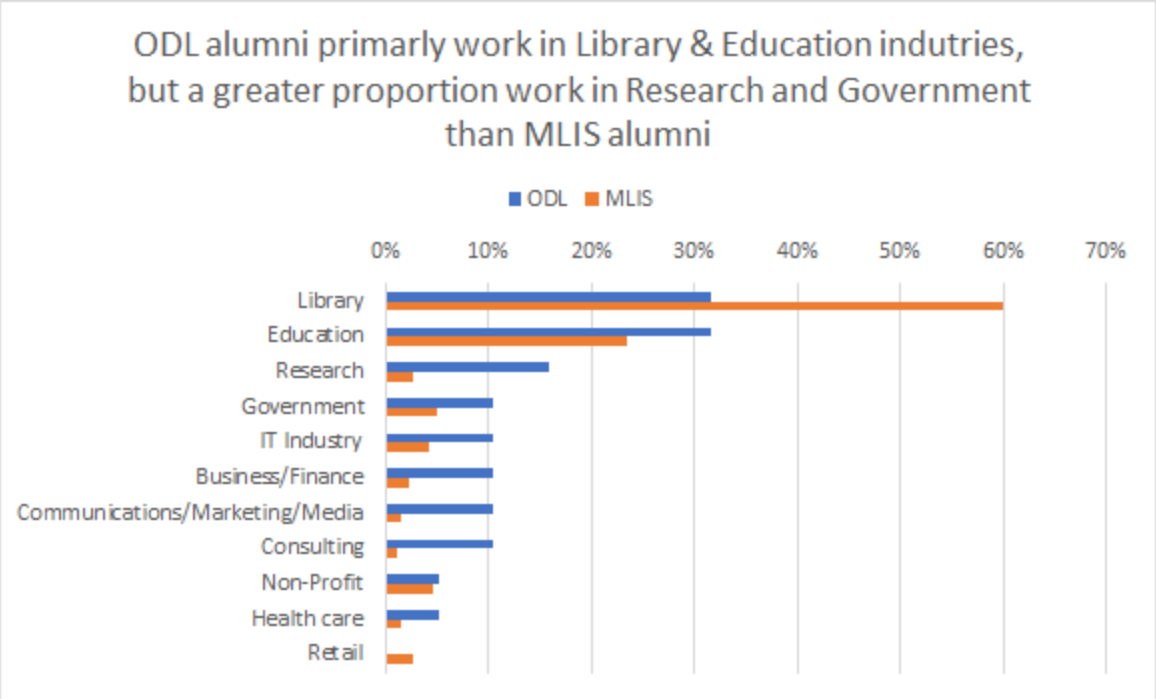


Figure 4.

ODL alumni work in a greater variety of industries. The third most common industry for ODL alumni is Research at 16%, but the third for MLIS only has 5% of respondents. ODL alumni reported a wide range of salaries: $20,000 - 109,999 a year. $70,000-79,999 was the mean salary.

Where librarians and library staff work

We have a particular interest in how librarians and library staff apply their data expertise at work. To better understand the kinds of data expertise they are using within libraries, we first identified all of the respondents that do library work. Upon reviewing respondents' job titles, we noticed that two did not list their industry as "library" but had "librarian" in their job title. We then paired all of the library job titles with the reported industries. (A full table of all job titles and industries is available in the Appendix.) Almost all of the positions classified their work as "library" and/or "education," and with the "Open Data Librarian" classifying their industry as "education" and "government."

Job titles	Library	Education	Government
Adult Services Librarian	x		
Cataloging and Project Intern	x		
Data and Digital Scholarship Librarian	x		

IT Business Analyst	x		
Library Systems Assistant	x	x	
Research and Instruction Librarian	x	x	
Scholarly Communication Librarian for Research Infrastructure		x	
Open Data Librarian		x	x

Using data expertise at work

Most of our respondents spend a majority of their time at work using their data expertise. We define data expertise as knowledge of and emerging skills using, managing, and analyzing data.

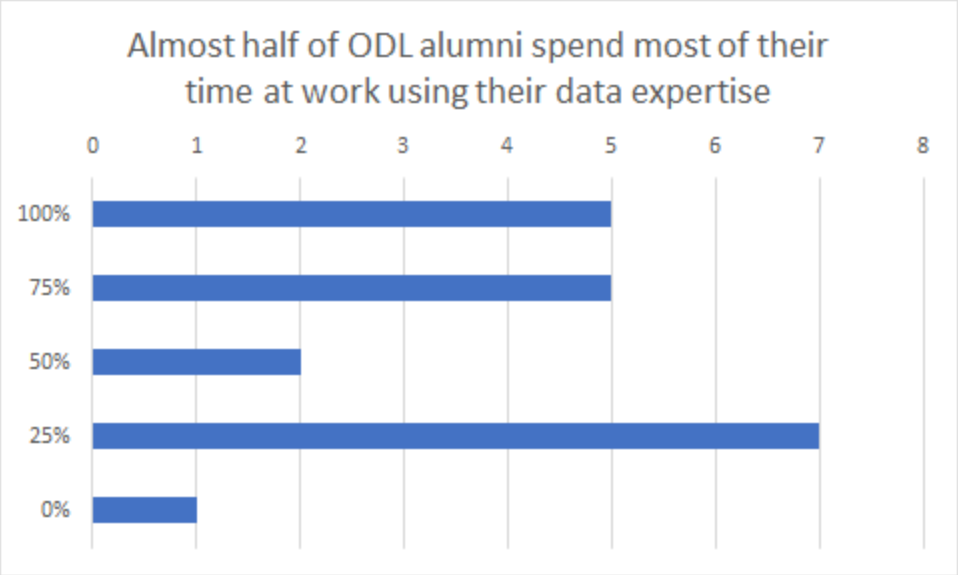


Figure 5.

In a follow-up question, ODL alumni shared that they use their data expertise in a variety of activities with the most common including data management, metadata and documentation, and user instruction and training.

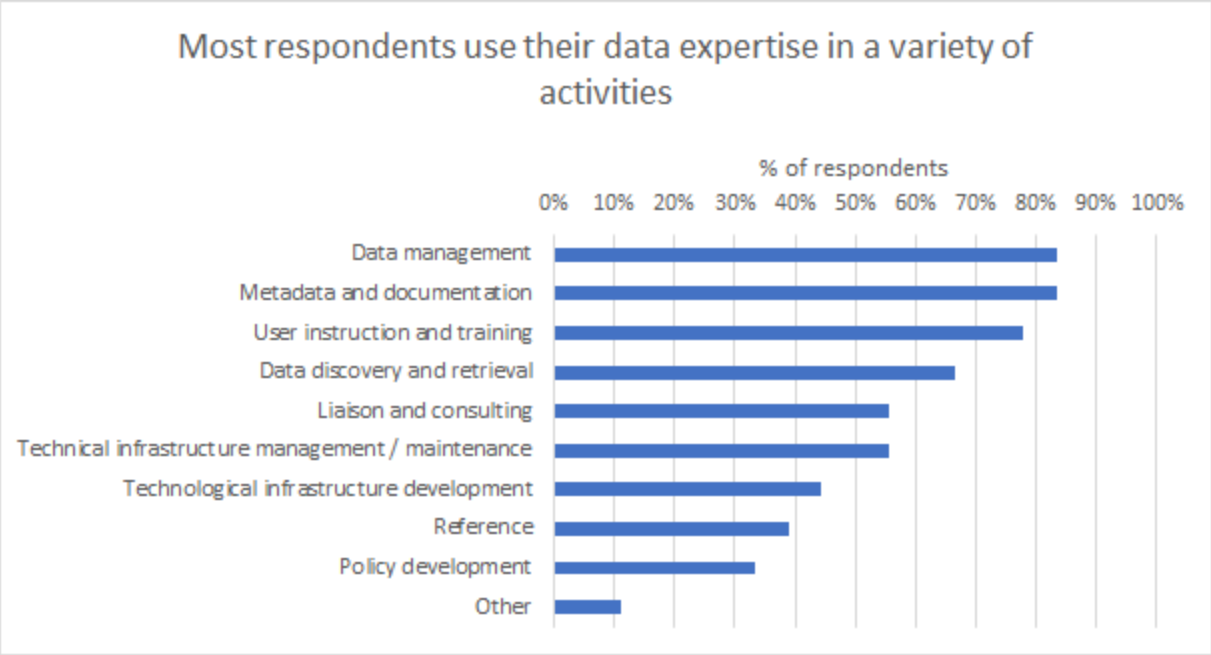


Figure 6.

In our open ended responses, some respondents gave more detail on how they use their data expertise in the context of these activities. A few respondents working in a library setting described managing bibliographic data to inform a “DEI [Diversity, Equity, and Inclusion] audit of our collections” and to inform “weeding decisions.” Another respondent working in research described deploying “multiple metadata schemas (custom, but built from community metadata standards and collective working group input) for a data repository.” Library-based respondents described user instruction and training such as “hosting data management workshops” for students, faculty, researchers and teaching a “Research Data Management course to [university] faculty.” These activities tend to be non-technical. However, about half of all respondents did do technical work: Respondents that engaged in technical activities worked in a variety of industries including libraries, universities, and private industry. Several respondents also reported using their data expertise in other categories such as data visualization, data literacy, data privacy, and writing and research.

Other open-ended responses had to do with how respondents were educating colleagues and advocating for data policy and practices. Although sometimes these interactions happened in more formal settings like workshops, they also happened in more ad hoc ways and as part of daily business. A library respondent said, “We are having a lot more conversations about data privacy and usage at my institution that were not happening before I joined.” Another valued bringing the “LIS perspective to the management of open government data.” A Product Specialist working in the IT Industry explained, “I use what I learned daily in communications with clients and colleagues. Understanding core concepts and terms allows me to comprehend details quickly and fill in knowledge gaps when needed.”

Curriculum

Next, we try to understand how the ODL curriculum and educational experiences, within the context of the UW Information School, impacted ODL alumni careers. We listed 13 areas of the curriculum and asked alumni to select all that contributed to their professional progress (Figure 7). Half or more of respondents selected 6 of these areas.

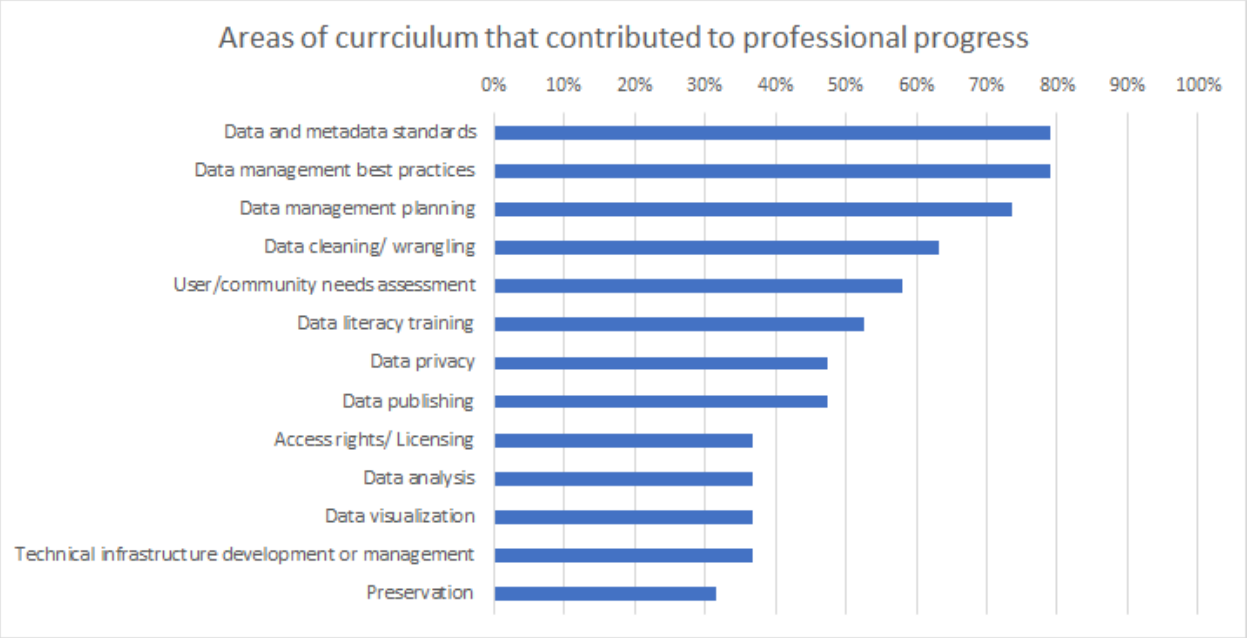


Figure 7.

The top three areas contributing to professional progress are data and metadata standards, data management best practices, and data management planning. These areas line up with the top activities that alumni apply their data expertise at work. In other words, alumni use what they learned from ODL directly in their careers. Fewer alumni selected the more technical aspects of the curriculum (data analysis, data visualization, technical infrastructure development) but 37% of alumni still found these areas useful in their career. A respondent currently employed in a library said, “these classes changed my focus trajectory.”

Alumni also report that what they learned at the iSchool prepared them to learn new data skills. 78% of respondents said that iSchool curriculum prepared them extremely well or very well.

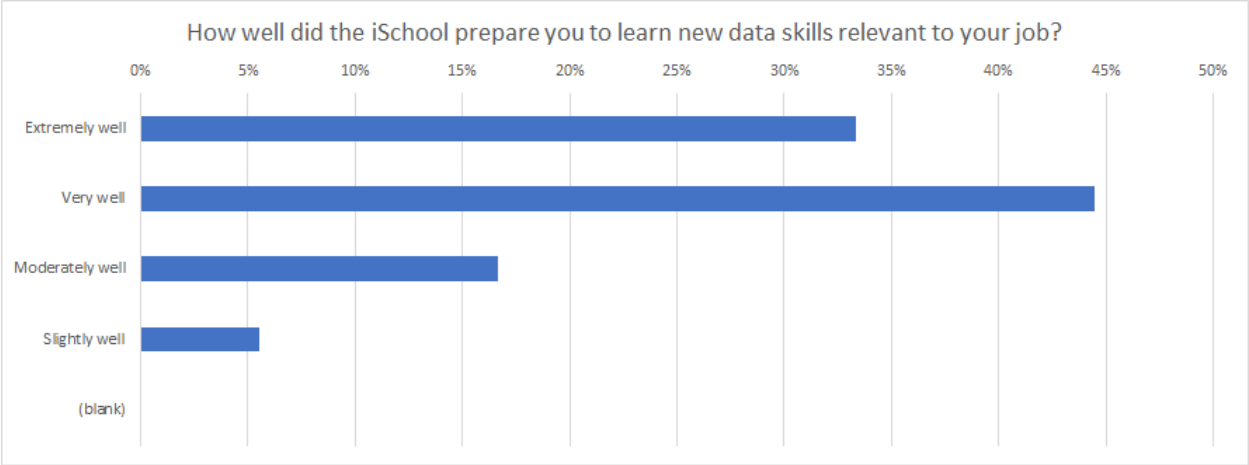


Figure 8.

A few individuals shared some insights on what made the curriculum successful for them. One respondent advised current and future students to “Make sure you spend enough time doing the assignments because you will really learn a lot from them!” Another noted: “creating a relationship with faculty members was very important to my understanding of the material.”

Someone working in private industry gave this advice to current and future students:

> *Understanding end-to-end flow of data and information is crucial. The data output is only a piece of the puzzle. Focus more on the input--how are we collecting data? Are we asking the right questions to measure what we want to learn about? What are our limitations?--and understand how data moves through a system. This will give you insight into how to interpret outputs and context to connect back to the knowledge gap you're trying to address.*

Suggestions for curriculum revision

We also asked alumni for suggestions on revising the curriculum. The broad area of data literacy was the most popular suggestion, but after that more technical topics ranked high (data analysis, data visualization, and technical infrastructure or management).

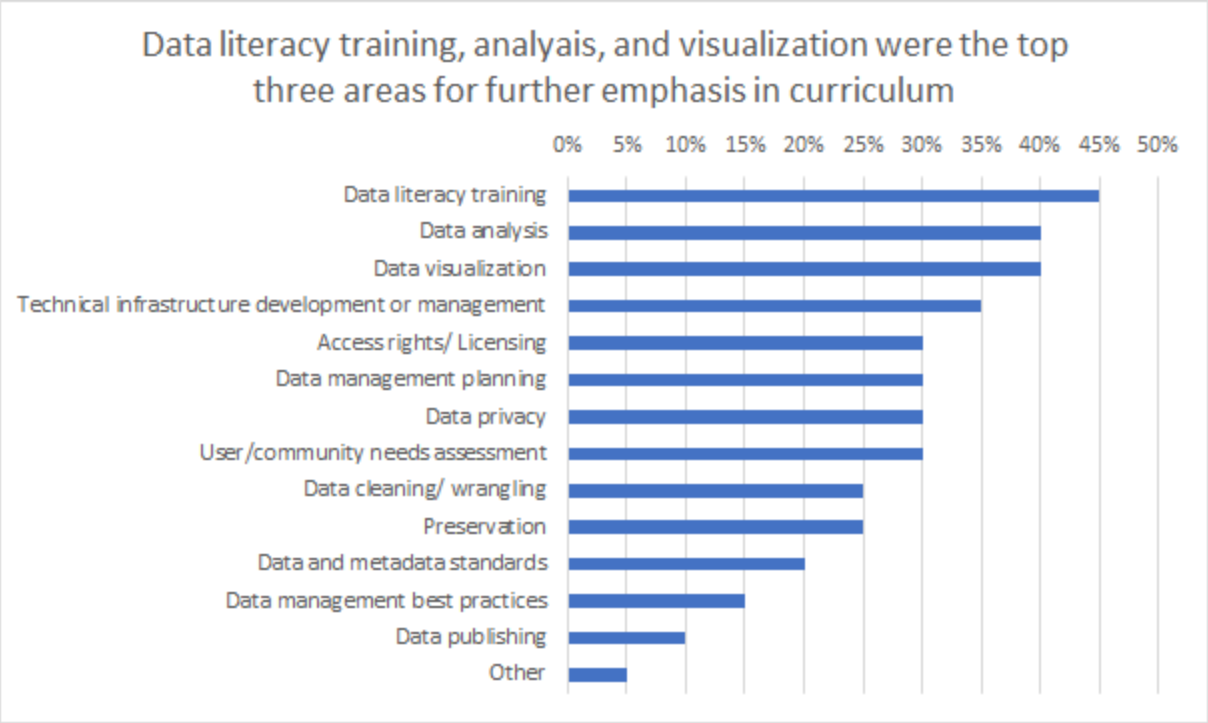


Figure 9.

For data literacy, a couple of respondents explained that they thought data literacy is valuable for all Library and Information Science students, particularly for the purpose of passing on that knowledge to others such as “public, organizations, customers, etc.” Another explained the need to “help patrons think critically about data and its role at work, in private life, in public life.” Others explicitly mentioned data ethics and “data collection’s influence on marginalized communities.” Another respondent recommended that a current iSchool “special topics” class on data ethics be made a permanent part of the curriculum. Several respondents mentioned data regulation and privacy such as HIPPA, FERPA, GDPR, and other international practices.

The high ranking of technical topics reflects an ongoing tension in coursework between a focus on the conceptual and the technical. On one hand, some alumni want more of a focus on the technical: “I have a great theoretical basis for doing this work but I don’t feel like I consistently got any training in the tech side.” On the other hand, other students have negative associations to more technical concepts which may steer them from taking the coursework: “students without a technical background were afraid to take some of the data-heavy courses… . Figuring out how to support students who do not have a technical background should be a focus.” One respondent summed up this conflict on a more personal level:

> *Balancing technical vs. conceptual understanding has been a challenge for me and I'm never sure how to best allocate my time. I didn't focus on technical coursework while at the iSchool (e.g. I did not take the data science courses or CS classes while at UW). This was a purposeful choice because I had technical expertise before attending the*

> *iSchool and felt confident in my ability to self-study. This tension is an ongoing challenge in my career; there's so much to learn from both the technical and conceptual perspectives and my focus at any given point in time changes based on the needs of my job & my interests.*

To be clear, respondents requesting more technical aspects in the curriculum were still the minority. Forty percent of respondents wanted more emphasis on data analysis and data visualization and 35% wanted emphasis on technical infrastructure development or management. However, it seems to be an unmet need for a sizable number of students and alumni.

A few respondents had some suggestions for preparation of data experts in relation to their organization: “I'd recommend students get some vocabulary for change management -- this is usually an issue when opening data or working with organizations on data management.” Another respondent, using slightly different language, agreed: “Advocacy is a really important part of open data...a concrete advocacy focus within the curriculum would be a nice addition.” One library respondent mentioned that “making open data a priority to leadership” is an obstacle to building “a presence for local data on the state open data portal.”

Discussion

It appears that ODL educational experiences did prepare students for professional positions and have assisted with upward mobility. Half of all 2017-2019 alumni have been promoted or changed jobs since graduation, and one 2020 graduate reported receiving a promotion 7 months after their hire date. Alumni also overwhelmingly reported that the curriculum supported them in learning new data skills. In examining the data, several themes emerged that will help inform future curriculum revisions.

Tension between technical and conceptual

As described in the previous section, there was a tension between the technical and conceptual aspects of the curriculum. This was revealed in formative evaluations and classes were iteratively updated to increase coverage of tools and technologies over time. However, many graduates experienced a more conceptual focus that nonetheless assisted in career growth, but there seems to be a sizable minority that needed (and still need) more technical instruction. Although some survey respondents gave some details on what specific technical skills they needed, follow-up conversation with alumni would be helpful in parsing out exactly which areas need more focus in the curriculum.

Communication/policy/change management

Communicating and advocating about open data and data management in general was a theme in the open ended responses. It appears that alumni may be the most knowledgeable or passionate about these topics within their organization and may have sole responsibility in these areas. Alumni may benefit from a professional cohort or network to continue to build expertise and support one another. RDAP (Research Data Access and Preservation Association) is one example of one such professional organization, but it is focused more on academic librarians and research data. There is some emerging research within the Information School on open data work at public libraries that could help inform best ways forward as well.

Limitations

The alumni survey had several limitations. First of all, due to iSchool procedure, we could not directly email all of the ODL alumni and had to use the iSchool Alumni relations office as an intermediary. Although the mailing did produce a high “engagement score” from the office’s software, we were concerned that alumni would be less likely to participate in the survey than if the email had come directly from the ODL faculty or team. In response, we sent a separate email from a faculty member’s email account to the subset of ODL internship alumni, which resulted in a disportionately high response rate from intern alumni. Additionally, 2020 alumni were more likely to respond to the survey which resulted in less information about longer-term career impacts. It is also possible that alumni that are more engaged with open data or actively using their data expertise were more likely to complete the survey. Alumni not working in these areas may not feel as much of a connection to the topic and be less motivated to respond.

We also attempted to use course evaluations as part of our analysis but found the amount of applicable information limited. For one thing, the closed-ended questions changed over the course of the three years of our project, and we could not look at comparisons over time. Most of the open-ended responses tended to focus on teaching style and method which was not the focus on this study.

Appendix 1A: Respondent Job Titles and Industries

Self-selected library industry or librarian job title.

Job titles	Library	Education	Government
Adult Services Librarian	x		
Cataloging and Project Intern	x		
Data and Digital Scholarship Librarian	x		
IT Business Analyst	x		
Library Systems Assistant	x	x	
Research and Instruction Librarian	x	x	
Scholarly Communication Librarian for Research Infrastructure		x	
Open Data Librarian		x	x

Typically for-profit industries and job titles

Job title	Business/ Finance	Communications /Marketing/Media	Consulting	Health care	IT Industry
Data Engineer	x		x		
Knowledge Management Advisor	x				
Media Asset Manager		x			
Senior People Insights Analyst		x			
Medical Receptionist				x	
Analyst			x		x
Product Specialist					x

Typically nonprofit industries and job titles

Job title	Education	Government	Nonprofit	Research	Unknown
Program Coordinator	x	x			
Bioinformatics Analyst, Data Curator			x	x	
PhD Student	x			x	
Research Specialist	x			x	
Process Analyst					x

Appendix 2:

Selected curricular materials

Information School – University of Washington

Data Curation 1: Fundamentals of Data Curation
LIS 545
Spring 2021

Instructor and TA Information

Carole L. Palmer
clpalmer@uw.edu
Office: 370M MGH

Office Hours: Tuesday 5:00-6:00
and by appointment

Jamie Ramos
reader/grader
jmramos@uw.edu

Course Description

Data curation is an emerging area of expertise for information professionals across sectors. This course introduces the principles and practices involved in the curation of digital research data to benefit research communities and organizations and open data to benefit the public. The course emphases curation expertise for development of accessible and reusable data resources, with a focus on data services in academic libraries and open access repositories to support research, as well as initiatives in government and public libraries to support open data for public use.

Required Course Materials

The course materials are provided as links or downloads in the Module pages.

Course Format/Structure

This course begins on Monday, March 29th. Each week covers one thematic area with readings, lecture, lab activities, and discussion. You will complete 3 individual assignments and 2-3 lab memos. You will also share concepts, thoughts, and observations on discussion forums. We will host a series of guest speakers in the middle of the term.
It is my hope that you will participate via video when you can, but you may elect to attend lectures and engage in class activities with audio only.

Note:
The course is scheduled to run synchronously for most of the scheduled class times via Zoom. These Zoom class sessions will be recorded. The recording will capture the presenter's audio, video, and computer screen. Student audio and video will be recorded if they share their computer audio and video during the recorded session. The recordings will only be accessible to students enrolled in course. These recordings will not be shared with or accessible to the public.

The University and Zoom have FERPA-compliant agreements in place to protect the security and privacy of UW Zoom accounts. Students who do not wish to be recorded should:
Change their Zoom screen name to hide any personal identifying information such as their name or UW Net ID, and not share their computer audio or video during their Zoom sessions.

Student Outcomes/Learning Goals

- Understand trends, principles, and practices in the emerging field of data curation. Comprehend data practices and curation requirements across disciplines and sectors.

- Build foundational knowledge of social and technical dimensions of developing and providing data curation services.
- Develop practical understanding of data curation work and the role of information professionals in the full lifecycle of data.

Student and Instructor Expectations

I expect graduate students to be adult learners and good citizens of the classroom. As a member of the classroom community, you are accountable to me, but, just as importantly, you are accountable to yourself and each other. Being a good citizen in this learning environment requires coming to class fully prepared. Without full preparation, you reduce the value of our collective learning and take undue advantage of the work of others. Additionally, a good citizen shares their observations, questions, experiences, and progress as well as their challenges and setbacks--all are highly productive contributions to our learning and class experience.

As your instructor, I am committed to your learning and your preparation as an information professional. I will always do my best to keep us on course, intellectually, and on schedule, in respect for your need to manage your time.

Assignments

I have designed the assignments and activities to complement and supplement the readings, lectures, and discussion.The three formal assignments will give you experience examining and analyzing data and metadata, and an opportunity to develop a proposal for a curation initiative for an organization. See the Assignments page for details and due dates.The Lab Activities and Resources, and associated short Memo assignments, are designed to build awareness and understanding of data curation resources, tools, and practical aspects of data services.

Grading

University policies on grading, academic conduct, and other general academic policies apply. See https://depts.washington.edu/infodocs/academic_policies.

All students are required to understand and abide by the Academic Policies of the Information School and the UW Student Conduct Code.

Note: Plagiarism is serious academic misconduct. In the Code, plagiarism is defined as:
the submission or presentation of someone else's words, composition, research, or expressed ideas, whether published or unpublished, without attribution. Plagiarism includes, but is not limited to the use, by paraphrase or direct quotation, of the published or unpublished work of another person without full and clear acknowledgment; or the unacknowledged use of materials prepared by another person or acquired from an entity engaging in the selling of term papers or other academic materials.

Please see this link for further details: http://www.washington.edu/admin/rules/policies/WAC/478-120-024.html

Late Assignments

Late assignments will be penalized, unless an extension has been granted prior to the due date.

MODULES

Module 1: Introduction to the Course, Definitions, and Concepts

Overview

This first week will serve as an introduction to the course. I'll talk through the weekly topics and the assignments to make sure you understand course expectations. We will also consider the concepts of "data" and "data curation" and related workforce trends for information professionals. Our first Lab focuses on the scope and complexity of data resources.

Objectives

- Grasp the scope of expectations for the course.
- Understand key concepts.
- Build awareness of workforce trends.
- Become familiar with the range and depth of data repositories and portals available to researchers and the general public.

Class Sessions

Monday, March 29**:**

Introduction to the course.

Prior to class, post a personal statement on the Introductions discussion forum, via video or text. Tell us a little about your background, academic and professional interests, and what drew you to this class.

Wednesday, March 31**:**

Lecture: Data Curation: Context & Definitions.

Assignment 1 overview.
Lab Discussion: defining data; landscape of data access
Jamboard: defining data

Readings

Required:

- Palmer, Carole L., Weber, Nicholas M., Muñoz, Trevor, & Renear, Allen H. (2013). Foundations of data curation: The pedagogy and practice of 'purposeful work' with research data. *Archive Journal, 3*.
- Borgman, C. L. (2015). What are data? In *Big Data, Little Data, No Data: Scholarship in the Networked World,* pp. 17-29. MIT press.

Trade press perspectives for public libraries and industry:

- Shueh, Jason. Data reinvents libraries for the 21st century. Government Technology, June 5, 2015.
- Knight, Michelle. (2017). Data curation 101: The what, why, and how. Dataversity: Data Education for Business and IT Professionals.

For Review:

- Owens, T. (2011). Defining data for humanists: Text, artifact, information or evidence? *Journal of Digital Humanities,* 1(1).
- What is research data? Australian National Data Service.
- What is open data? Open Data Handbook. Open Knowledge International.
- Selected organizations:
  - Research Data Alliance (RDA)
  - Digital Curation Centre (DCC)

- FORCE11. Future of Research Communication and e-Scholarship
- SPARC. On Open Data.
- Code4Lib
- Committee on Data for Science and Technology (CODATA)
- World Data System (WDS)

Lab Activity & Resources

1. Begin a Data Curation Lab Journal for documenting your lab activities and other class learning. This is for your own personal use and will not be graded, but note--some Labs have associated homework assignments or class discussions, as is the case this week with our Wednesday session.
2. Write down a definition of "data" that you think works within the context of "curation" and the provision of data services to user communities.
3. Explore the landscape of data repositories and portals available through the access sites below. Make observations in your journal about how easy or hard it is to get to a dataset, download it, and understand the content.

Global registry of research data repositories:
- re3data.org

U.S. open government data**:**
- data.gov [limited to State of Washington]
- Washington State Open Data. Catalog
- data.seattle.gov

Module 2: Underpinnings of Data Sharing and Access Across Disciplines

Overview

This week is grounded in foundational reading providing different disciplinary perspectives on data curation and important principles and policies that have evolved to help guide data practices and services. My lecture covers important concepts and models and provides background on cyberinfrastructure, publishing, archiving, openness, and sharing. The Lab introduces you to data dictionaries, an essential data documentation practice.

CONCEPT OF THE WEEK forum posting - Due April 5
LAB MEMO - Data Dictionaries - Due April 7

Objectives

- Understand key concepts and different disciplinary perspectives and priorities for data curation and data services.
- Build awareness of high impact efforts to form policies and principles for data management, sharing, and curation.
- Learn about and discuss the value of data documentation in the form of data dictionaries.

Class Sessions

Monday, April 5:
Lecture: Data Sharing and Access across Disciplines.

Wednesday, April 7:
Discussion: Concepts - Jamboard

Readings

Required:

- Gray, J., Szalay, A. S., Thakar, A. R., Stoughton, C., & vandenBerg, J. (2002). Online scientific data curation, publication, and archiving. Microsoft Research Technical Report. Redmond, WA: Microsoft Research.
- Gutmann, M., Schürer, K., Donakowski, D., & Beedham, H. (2004). The selection, appraisal, and retention of social science data. *Data Science Journal*, *3*(30), 209–221.
- Munoz, T. (2013). Data curation as publishing for the digital humanities. *Journal of Digital Humanities*, 2(3).
- NYU Health Sciences Library. Data Sharing, Part 1. Request. YouTube video.

For review**:**

- White House memos:
  - Expanding Public Access to the Results of Federally Funded Research. White House, Office of Science and Technology Policy (OSTP), February, 2013.
  - Open Data Policy. White House, Executive Order, May, 2013.
  - Request for Information: Public Access to Peer-Reviewed Scholarly Publications, Data and Code Resulting From Federally Funded Research. White House, OSTP, February 2020.
- FAIR guiding principles
  - See also: Mons, Barend, et al. (2017). Cloudy, increasingly FAIR; Revisiting the FAIR data guiding principles for the European Open Science Cloud. *Information Services & Use* 37: 49–56.
- Open letter to PLos: Libraries role in data curation .
- Nature. Scientific Data. Recommended Data Repositories.

Background on cyberinfrastructure:

- Atkins, D. (2003). Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure. National Science Foundation.
- American Council of Learned Societies (ACLS). Commission on Cyberinfrastructure for the Humanities and Social Sciences. (2006). Our Cultural Commonwealth. New York, NY: ACLS.
- Edwards, P. N., Jackson, S. J., Bowker, G. C., & Knobel, C. P. (2007). Report of a workshop on "history & theory of infrastructure: lessons for new scientific cyberinfrastructures". Understanding Infrastructure: Dynamics, Tensions, and Designs.

Lab Activity and resources

Read Data dictionaries by Kristin Briney (Ab Initio, 2014). Look at the original files she used to demonstrate the data dictionary. Download and inspect the Data Record and the Readme file from the Dryad repository at: https://datadryad.org/resource/doi:10.5061/dryad.fj974 .

Other resources for review:

- Describing your data: Data dictionaries. Smithsonian Data Management Best Practices, 2018.
- How to make a data dictionary. Open Science Framework Guides.
- USGS. Data Dictionaries.
- Collection of data dictionaries associated with the National Institute of Mental Health Data Archive at NIMH Data Archive. Data Dictionary: Data Structures.

Lab Memo Assignment: Write a short memo (about a paragraph or up to about 150 words) on the purpose and value of data dictionaries and how they might differ in content and use across disciplines. Feel free to share any interesting or instructive examples.

Module 3: Introduction to Metadata for Data

Overview

Metadata is essential to supporting both access and use of data, but there are many challenges in application. This week is focused on the characteristics of metadata for data and the growing array of applicable standards. We consider the purpose of metadata, types, and levels of description, and examine selected standards and best practices.

ASSIGNMENT 1 - REPOSITORY EXPLORATION & COMPARISON, due Friday, April 16

Objectives

- Understand the central role and value of metadata in data curation.
- Become familiar with common structures, standards, and best practices, and the role of machine readable metadata.
- Through the Lab Resources, gain exposure to APIs.

Class sessions

Monday, April 12:

Lecture - Metadata for Research Data

Wednesday, April 14:

Lecture - Metadata Part 2: Levels of Granularity.
Assignment Q & A

Readings

Required:

- Getting Meta with Metadata: A City's Guide to High Quality, Discoverable, and Understandable Open Data, pp. 3-7. Bloomberg Philanthropies.
- Riley, Jenn. (2017). Types of Metadata. In *Understanding Metadata*. NISO, pp. 6-7.
- ICPSR. Best Practices for Creating Metadata, pp. 32-36.
- W3C. Data on the Web Best Practices. 8.2 Metadata. January 31, 2017.
- Farnel, S., & Shiri, A. (2014). Metadata for research data: Current practices and trends. *Proceedings of International Conference on Dublin Core and Metadata Applications*. Dublin Core Metadata Initiative.
- Cornell University. Research Data Management Service Group. Guide to Writing Readme Style Metadata.

Strongly Recommended for Review:

Metadata resources

- DataOne. Metadata. [Guidance for scientists on metadata]
- Data Documentation Initiative (DDI)
- Research Data Alliance. Metadata Directory
- Fairsharing.org. Standards.
- GovEx Labs. Civic data standards
- W3C. Data Catalog Vocabulary (DCAT)
- Open Data Institute. Marking up your dataset with DCAT

Lab resources

- Norlander, Bree. How to use an API (and what is it anyway?). API Lecture 720p (1).mp4
- Tutorial**:** Intro to API for open government data. (by An Yan and Bree Norlander).

Module 4: Open Government Data

Overview

As the open data movement grows, government data has become a central focus for empowering the public and increasing government transparency. This module provides background on open data for public use and the emerging role for public libraries in the open data ecosystem. The Lab introduces resources for conducting a data inventory, a process of growing importance for city open data initiatives.

LAB MEMO - Data Inventories, Due April 21

Objectives

- Understand trends in access to open government data in relation to research data.
- Consider benefits, principles and policy guidelines for open data access.
- Relate open data activities and trends to the present and future roles of information professionals and libraries.
- Assess the value and work involved in conducting a data inventory for a city or organization.

Lecture

Monday, April 19:

Lecture: Open Government Data.

Wednesday, April 21:

- Guest Speaker: Kathleen Sullivan, Open Data Literacy Consultant, Washington State Library. Open Data and Public Libraries.
- Assignment 2 overview

Readings

Required:

- Okamoto, Karen. (2017). Introducing open government data. *The Reference Librarian* 58(2).
- Robinson, Pamela, and Mather, Lisa Ward. (2017). Open Data Community Maturity: Libraries as Civic Infomediaries. *Journal of the Urban and Regional Information Systems Association*, 28.
  - [See pages 31-38 of special issue on infomediaries, and review other articles related to open data.]
- Yoon, A., Copeland, A., & McNally, P.J. (2018). Empowering communities with data: Role of data intermediaries for communities' data utilization. 81st Annual Meeting of the Association for Information Science and Technology (pp. 583-592). Somerset, NJ: Wiley.
- Kathleen Sullivan. (2018). Open Data and Libraries. Washington State Library, First Tuesdays Webinar series.

For Review**:**

- Horrigan, John B., and Rainie, Lee. (2015). Americans' Views on Open Government Data. Pew Research Center.
- Open Data Handbook. Open Knowledge Foundation.
- Responding to Open Data Concerns. GovEx Labs, Center for Government Excellence, Johns Hopkins University.
- Open City. Civic Apps Built with Open Data.
- Open Data: Guidance: What Data to Publish. Washington State, Office of the Chief Information Officer.

Lab activity and resources

1. Read "How to Create a Dataset Inventory" (pages 3-7) in the City of Seattle Open Data Playbook. Seattle Information Technology, 2016.
2. Review the City of New Orleans, Building Your Data Inventory materials. Download the example Safety and Permits inventory.

Additional resources on city data inventories, for consultation:

GovEx Labs. Data Inventory Guide. Johns Hopkins University.
DataSF. Data Coordinator Guidance. City of San Francisco.

Lab Memo. Write a short memo (about a paragraph or up to about 150 words) on:

a) What are the strengths of the New Orleans inventory approach. What could be enhanced or improved?
b) What challenges would you expect to encounter carrying out such an inventory?

Module 5: Data Identity, Context, and Process

Overview

The capture, representation, and presentation of data in relation to its original and evolving context is key to meeting the goals of data curation. In this module we examine the concepts and activities involved in retaining and propagating the context needed for data to be successfully managed, transferred, shared, and reused.

CONCEPT OF THE WEEK forum posting – Due April 26

Objectives

- Develop a base of knowledge on the identity and context of data objects and how they are are managed over time.
- Become familiar with curation practices associated with provenance and attribution.
- Distinguish research and data production workflows from curation workflows.
- Relate data context concepts and activities to metadata, quality control, reuse, and trust.

Class sessions

Monday, April 26

Lecture: Data Identity, Context, Process.

Wednesday, April 28

Guest Speaker: Kaitlin Throgmorton, Sage Bionetworks. Demystifying Biomedical (Meta)Data.

Readings

Required:

- Viglas, S.D., (2013). Data provenance and trust. *Data Science Journal,* 12, pp.GRDI58–GRDI64.
- Data Citation Synthesis Group. Joint Declaration of Data Citation Principles. (2014). Martone M. (ed.), San Diego CA: FORCE11.
- Groth, P., Cousijn, H., Clark, T., & Goble, C. (2020). FAIR data reuse – the path through data citation. *Data Intelligence*, 2(1-2), 78-86.
- Australian National Data Service. Persistent Identifiers: Working Level.
- Australian National Data Services. Data Provenance.

For Review:

- Goodman, Alyssa, et al. 10 Simple Rules for the Care and Feeding of Scientific Data. *PloS Computational Biology* 10(4): e1003542.
- Hills, D.J., Downs, R. R., Duerr, R., Goldstein, J. C., Parsons, M. A. & Ramapriyan, H. K. (2015). The importance of data set provenance for science. *Earth & Space Science News*, 4 December.
- Thomer, Andrea, K., Wickett, Karen M., Baker, Karen S., Fouke, Bruce W., and Palmer, Carole L. (2018).
  Documenting provenance in noncomputational workflows. *Journal of the Association for Information Science and Technology*, 69(10), 1234-1245.
  Crotty, David. (2014). Reproducible research: A cautionary tale. The Scholarly Kitchen. Blogpost. 26 March 2014. [Read article and review comments.]
- DataCite.
  - Cite your data.
  - Format your citation.
- Renear, A.H., Dolan, M., Trainor, K., Cragin, M.H. (2009). Towards a cross-disciplinary notion of data level in data curation. *Proceedings of the 72nd ASIS&T Annual Meeting.* Vancouver, BC.
- Goble, C., Stevens, R., Hull, D., Wolstencroft, K., & Lopez, R. (2008). Data curation + process curation=data integration + science. *Briefings in Bioinformatics*, *9*(6), 506–517. Doi:10.1093/bib/bbn034

Lab Resources

Open Refine is a tool for data cleaning and normalization. Library Carpentry, listed first below, has the best overall coverage designed for librarians. Exploration and tinkering is highly recommended. You can use your API knowledge from Module 3 to access some metadata for experimentation, as demonstrated in the tutorial below.

- Library Carpentry. Open Refine
- Open Refine
- Cleaning Data with Open Refine. The Programming Historian.
- Tutorial: Working with civic data portal metadata with OpenRefine (by An Yan)

Module 6: Data Services

Overview

Services to support data sharing and reuse are provided by repositories, research centers, libraries, and government. In this module we examine service models and levels of service get a practice based service perspective from a practicing data curation expert.

LAB MEMO – DMPTool, Due May 3

Objectives

- Gain understanding of the different investments and approaches to data services for academic libraries, domain repositories, and open data portals.
- Build awareness of growing expectations and changing priorities in data services.
- Learn first-hand from a case of applying open data to inform public library services.
- Examine approaches and tools for supporting data management work and meeting data sharing expectations.

Class session

Monday, May 3

Lecture - Data Services

Wednesday, May 5

Guest Speaker – Karalyn Ostler - Pubic Library Data Dashboard

Readings

Required:

- Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Baird, L., Sandusky, R., Langseth, M. & Lundeen, A. (2015). Research data services in academic libraries: Data intensive roles for the future? *Journal of eScience Librarianship* 4(2): e1085.
- Lyle, Jared. (2017). Making the case for disciplinary data repositories. In Curating Research Data, Volume II: A Handbook of Current Practice, by Lisa R. Johnston. ACRL, pp. 162-164.
- Ostler, K. R., Norlander, B., & Weber, N. (2020). Using open data to inform public library branch services, *Public Library Quarterly*, DOI: 10.1080/01616846.2020.1798206
- Sieber, R. E., & Johnson, P.A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, 32(3): 308-315.

For Review:

- Research Data Alliance. 23 Things: Libraries for Research Data.
- Fallaw, Colleen, et al. (2016). Overly honest data repository development. *Code4Lib Journal*, Issue 34.
- *Bulletin of the Association for Information Science and Technology*, 31(6). Special section: Research Data Access and Preservation (RDAP) Summit 2015.
- Lee, M., Almirall, E., & Wareham, J. (2016). Open data and civic apps: First-generation failures, second-generation improvements. *Communications of the ACM*, 59(1): 82-89.
- Police Data Initiative. Not just for books: Public libraries partnering with police to engage communities with open data.

Lab activity & Resources

Review the example DMPs provided by the University of Minnesota Libraries.
Set up a DMPTool account using your UW credentials.

1. Review a few of the UW shared plans listed in the dashboard.
2. Compare two templates by starting a couple of TEST plans. For example, set up a plan for a foundation grant, such as the Gates Foundation, and one for a National Science Foundation grant within one of the divisions, such as the Social, Behavioral, Economic Sciences.

Lab Memo: Write a short memo (about a paragraph or up to 150 words), based on your observations from reviewing the UM examples, the UW shared plans, and the DMP templates and tool functionality, more generally.

How self sufficient do you think researchers can be using the DMPTool?
What additional guidance or support, if any, would be most valuable for researchers in creating high quality data management plans?

Module 7: Policy, Rights, Privacy, Ethics

Overview

Providing access to data requires a broad range of policy and ethical considerations. This module offers an introduction to the many facets where policy comes into play. We will examine issues of copyright, licensing, and means for protecting privacy of human subjects data and other ethical concerns, including indigenous data sovereignty.

Assignment 2 – Metadata Comparison, Due May 10

Objectives

- Comprehend the diverse policy landscape that impacts the practice of data curation and provision and use of open data.
- Build awareness of ethical expectations and responsibilities for practicing data curation professionals.
- Become familiar with the application of policies in open data portals and research data repositories.
- Learn from a current case of ongoing develop of data services for qualitative data, with a focus on de-identification.

Class Sessions

Monday, May 10

Lecture - Policy, Rights, Privacy, and Ethics

Wednesday, May 12

Guest Speaker. Sebastian Karcher , Associate Director, Qualitative Data Repository. Curation of Qualitative Social Science Data--focus on de-identification.

Readings

Required:

- Locke, B & Weber, N (forthcoming). Ethics of open data. In *Scholarly Communication Librarianship and Open Culture: Law, Economics, and Publishing.* ACRL.
- Carroll M. W. (2015) Sharing Research Data and Intellectual Property Law: A Primer. *PLoS Biology* 13(8): e1002235.
- Center for Open Data Enterprise. (2016). Briefing Paper on Open Data and Privacy.
- Open Data Policy Hub. Sunlight Foundation.
- Indigenous Data Sovereignty. Data for Governance: Governance for Data. Briefing paper, 2018.

For Review:

- Creative Commons. CC0.
- Australian National Data Service. Publishing and sharing sensitive data.
- Australian National Data Service. De-identification.
- ICPSR. Confidentiality.
- Elman, C., & Kapiszewski, D. (2014). Data access and research transparency in the qualitative tradition. *Political Science and Politics*, January 2014.
- Letter from distinguished political scientists urging nuanced journal interpretation of JETS policy guidelines. *PSNow*, January 13, 2016.
- Stodden, V. (2009). The legal framework for reproducible scientific research: Licensing and copyright. *Computing in Science & Engineering, 11*(1), 35-40.
- Mellor, David. (2018). The landscape of open data policies. Center for Open Science.

- Mannheimer, Sara, and Hull, Elizabeth A. (2017). Sharing selves: Developing an ethical framework for sharing social media data. *International Journal of Digital Curation*, 12(2), 196-209.
- EU's General Data Protection Regulation (GDPR)

Lab Resources

The GitHub software development platform is used by many teams and organization for coordinating work. For this Lab, consider how it be applied in curatorial processes.

- GitHub
- Library Carpentry. Intro to Git.
- Mozilla Science Labs. A Friendly Github Intro Workshop.

Module 8: Preservation and Rescue

Overview

The preservation of digital data is a fundamental curation responsibility for many organizations. It requires active management to reduce threats to long-term value, mitigate risk of obsolescence, and ensure access and the ability to reuse data over time. This module provides and introduction to preservation issues and practices related to policies, processes, and resources, and the relationship of data rescue and refuge efforts in libraries and the sciences.

CONCEPT OF THE WEEK forum posting - Due May 17
LAB MEMO - Preservation collaboration, Due May 19

Objectives

- Learn about digital preservation concepts and strategies for application in data curation.
- Obtain an introduction to Open Archival Information System (OAIS) model, an important foundation for preservation in data repositories.
- Gain exposure to preservation metadata, certification criteria, and tools to support preservation work.
- Consider the role of data rescue and similar efforts in the data ecosystem.

Class Sessions

May 17
Lecture - Preservation

May 19
Assignment 3 Overview
Discussion

Readings

Required:

- What is Digital Preservation. Consortium of European Social Science Data Archives.
- Phillips, M. (2013). The NDSA levels of digital preservation: An explanation and uses. Proceedings of the Archiving (IS&T) Conference, April 2013, Washington, DC.
- Navale, V., & McAuliffe, M. (2018). Long-term preservation of biomedical research data. *F1000Research*, *7*, 1353. doi:10.12688/f1000research.16015.1
- Hart, E. M., Barmby, P., LeBauer, D., Michonneau, F., Mount, S., Mulrooney, P., ... & Hollister,

J. W. (2016). Ten simple rules for digital data storage. *PLoS Computational Biology*, *12*(10), e1005097.

- Downs, R. R., and Chen, R. S. (2017). Curation of scientific data at risk of loss: Data rescue and dissemination. In *Curating Research Data: Practical Strategies for Your Digital Repository*, pp. 263-277. Edited by Lisa Johnston. Chicago, IL: Association of College and Research Libraries.
- Selected papers in special issue of GeoResJ: Rescuing Legacy Data for Future Science.
    - Lehnert, K., Hsu, L., Parsons, M. A., & Wyborn, L. (2015). Rescuing legacy data for the future of science. *GeoResJ* 6, 106-107.
    - Griffin, R. E. (2015). When are old data new data? *GeoResJ* 6, 92-97

For Review:

- OAIS Reference Model materials: Lavoie, Brian. (2014).
- The Open Archival Information System (OAIS) Reference Model: Introductory Guide. 2nd edition. DPC Technology Watch Series. Digital Preservation Coalition.
- Lee, C. A. (2009). Open Archival Information System (OAIS) Reference Model. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of Library and Information Sciences*, Third Edition, (pp. 4020-4030). Boca Raton, FL: CRC Press.
- See OAIS reference model document here.
- ICPSR Digital Preservation Policy Framework.
- Data Appraisal and Ingest. Consortium of European Social Science Data Archives.
- Core Trust Seal.
- Allen, L., Wright, Stewart, C., & Wright, S. (2017). Strategic Open Data Preservation. *College and Research Library News*, 78(9).
- Federal Data Refuge Initiatives & Tools. Portland State University Library.

Lab Activity and Resources

Review the array of tools, outlined in the POWRR grid and itemized in the COPTR tool registry. Assess the practical value of the open solutions wiki managed by the UK digital preservation community.

- POWRR tool grid. DigiPres Commons.
- COPTR. Community Owned Digital Preservation Tool Registry
- Digital Preservation and Data Curation Requirements and Solutions. Open Preservation Foundation.

Lab Memo: Write a short memo (about a paragraph or up to 150 words) that responds to the following scenario:

> You are an information professional in a university library or government agency. In your role, you are responsible for the curation and preservation of data resources but dependent on the institutional IT unit for infrastructure and technology solutions.
>
> What limitations or misconceptions might IT personnel have regarding preservation? How would you collaborate with them to evolve a robust preservation system?

Module 9: Sustainability and Costs

Overview

While the value of open data is often assumed, there is limited understanding of the costs involved in making data accessible and usable, and business models for sustaining data services are still evolving. This module offers background on the challenges in developing sustainable data services and models for determining costs.

Objectives

- Understand costs of data services in relation to specific curation activities and in relation to the lifecycle of data stewardship.
- Develop a base of knowledge on the current funding and business models for sustaining data services.
- Gain practical understanding of the cost / benefit tradeoffs and realistic value propositions for open data initiatives and research data services.

Lectures

May 24

Lecture - Sustainability and Costs

May 26

Discussion

Readings

Required:

- Johnson P. A., Sieber R. E., Scassa T., Stephens M., Robinson P. J.. (2017). The cost(s) of geospatial open data. *Transactions in GIS*, 21, 434–445.
- Erway, R., and Rinehart, A. (2016). If You Build It, Will They Fund? Making Research Data Management Sustainable. Dublin, Ohio: OCLC Research.
- Benison, Michael. (2016). Open data - how much does it cost? GovEx.
- Davidson, J. (2014). Investing in Curation: A Shared Path to Sustainability. Discussion Paper. 4C Project.

For Review:

- Berman, F., & Cerf, V. (2013). Who will pay for public access to research data? *Science*, *341*(6146), 616–617. doi:10.1126/science.1241625
- OECD. (2017). Business Models for Sustainable Research Data Repositories. OECD Science, Technology, and Innovation Policy Papers, no. 47.
- The Royal Society Science Policy Centre. (2012). Science as an open enterprise. "Costs of digital repositories". Blog post.
- Curation Costs Exchange.
- Palaiologk, A. S., Economides, A. A., Tjalsma, H. D., and Sesink, L. B. (2012). An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS. *International Journal of Digital Libraries*, 12, 195-214.

Lab Resources

These resources are designed to guide estimation and assessment of curation and data management costs. You will find them relevant to your final assignment, as you develop a plan for implementing your proposed curation service activity or upgrade.

UK Data Service. Data Management Costing Tool and Checklist

Costs of Data Management . Utrecht University.

Questions for consideration:

- How self-sufficient do you think researchers and other data producers can be in estimating their data costs over time?
- Why should data curators and data services operations invest in helping data producers understand and calculate costs?

Module 10: Review and Synthesis

Overview

This week is reserved for wrap-up of the course content and review.

Assignment 3. Curation Initiative Proposal Presentation. Due June 7

Class Session

June 2 - Review lecture and discussion

ASSIGNMENTS

Assignment 1
Repository Exploration & Comparison - 20 points, Due April 16

Objectives:

- Become familiar with different types of openly available data and associated repository contexts: city data in open data portals and research data in the context of domain, institutional, and multi-disciplinary repositories
- Gain experience analyzing the characteristics of datasets and repository service environments.

Preliminary exploration:

Tour the 9 repository sites listed below. Where indicated, examine the specific featured components.

City data portal examples:

1. Surrey's Open Data
2. Chicago Data Portal. Featured component: Dataset record: Mural Registry

Research data repository examples:

Domain repositories

1. UK Data Archive (social sciences)
2. tDAR (archaeology)
3. National Snow and Ice Data Center (geosciences). Featured component: NSIDC Data Policies

Institutional data repositories

1. University of Michigan, Deep Blue Data. Featured component: Depositor's guide.
2. University of Minnesota, Data Repository for U of M. (DRUM).
   Featured component 1: Policies and Terms of Use
   Featured component 2: Dataset record, particularly files at the bottom: An Analysis of Microlitter and Microplastics

Multi-disciplinary repositories

1. figShare
2. Zenodo
   - Featured component 1: Policies
   - Featured component 2: Dataset record.

Write an analysis of the two cases outlined below, guided by the prompts provided.

A. Zenodo case. Download and consult the dataset file.

Describe the data product.

- *Cover aspects such as the file type, size, names of creator(s), title, content, and structure.*
- *What can you discern about the original purpose or application?*
- *What are your impressions of the quality of the dataset?*
- *What is the potential for reuse of the dataset? Consider aspects such as scope of the content, user communities that might find it of interest, and other content and technical factors that may help or hinder application. Note difficulties assessing value for reuse*.

B. DRUM case. Download and consult the Readme file at the bottom of the record for the microlitter study. Then review the repository site more broadly.

Describe the value added by the content in the readme document.

- *What further understanding does it provide about the data? About the study?*

Comment on the UM libraries as a hosting institution.

- *How does hosting data of this kind align with their mission and roles?*
- *How does that alignment compare to cities that host data portals and that of other types of research data repositories.*

C. Conclusion:
Share a few selective, overarching observations about the range and variation in data products and the data services provided, considering the cases and the range of sites explored.

- *Which might be considered the least curated vs. the most curated of the examples? Why?*
- *Which seem to excel in curation and services in terms of ease of deposit? Access? Reuse of data? Why?*

Your paper does not need to be comprehensive. Focus on the learning objectives and what you determine to be the most important points in relation to the prompts of the assignment.
Note: Do not focus on interface design. Do not elaborate on metadata or catalog information, since this will be the objective of Assignment 2.

Deliverable:
Short, formal paper. Maximum of approximately 750-1000 words, single spacing preferred. With all assignments, construct the text as if you are addressing peer data curators who have a professional interest in your analysis. Use APA style for references to any sources or documentation reviewed. Submit the paper as a Word document (.docx) under 'Assignments' on Canvas. Please use the following file naming convention: [last name]_A[assignment #]_DC1_Sp21.docx (e.g. Smith_A1_DC1_Sp21.docx).

Grading rubric
- Evidence of detailed examination of the repository sites, featured components, downloaded data and records.
- Careful description and thoughtful discussion in line with the assignment prompts.
- Evidence of understanding of how repository environments and their data resources vary.
- Overall coverage and quality of analysis.
- Clear, professional narrative.

Assignment 2

Metadata Comparison – 15 points, Due May 10

Objectives:

- Become familiar with the metadata and documentation approaches used by various kinds of repositories.
- Consider implications for discovery, reuse, and interoperability.

A. Dataset Metadata Records and Documentation.

Compare the metadata records in the examples linked below for ICPSR and Dryad.

- ICPSR: Study of Jazz Artists, 2001 (Links to an external site.).
- Dryad: Data from: The unique ecology of human predators (Links to an external site.)
    - *Identify key differences in what is described, and how is it is presented, in the records provided.*
    - *What can you determine about the metadata schemes or standards applied?*
    - *What additional documentation is provided with the data files? What role do these items play relative to the records.*

B. Metadata Guidance.

Compare the metadata guidance offered by ICPSR, Dryad, and the Chicago Data Portal.

- ICPSR - Best Practices for Creating Metadata (Links to an external site.)
- Dryad (Links to an external site.) – "comprehensive metadata" link, through the "upload" link in the "Submit" section.
    - *What are the key differences in these two approaches to metadata guidance?*
    - What factors might explain the differences? Audience, scope of collection, range of service communities?
- Chicago Data Portal (Links to an external site.). Go to the "API documentation" provided by the software platform, Socrata, via the link at the very bottom of the page. That will take you to "Building something using data", then navigate to the "Metadata" option.
    - *Explain the difference in emphasis for this type of guidance, for technology developers rather than data depositors.*
    - *What might developers learn from ICPSR Best Practices or data depositors learn from developer guidance*?

C. Conclusion.

Share your overarching observations about the range and variation in metadata approaches and the implications for discovery, reuse, and interoperability. Consider the tradeoffs in producing rich vs. lean metadata and the return on investment for value added curation, such as links to related materials and representation of different levels of granularity.

Your paper does not need to be comprehensive. Focus on the learning objectives and what you determine to be the most important points in relation to the prompts of the assignment.

Deliverable:

Written assignment, with maximum of approximately 750 words, single spacing preferred. With all assignments, construct the text as if you are addressing peer data curators who have a professional interest in your analysis. Use APA style for references to any sources or documentation reviewed.

Submit the paper as a Word document (.docx) under 'Assignments' on Canvas. Please use the following file naming convention: [last name]_A[assignment #]_DC_Sp21.docx (e.g. Smith_A2_DC_Sp21.docx).

Grading rubric:

- Evidence of in-depth examination of the metadata and supporting documentation
- Evidence of understanding of metadata schemes, encoding, and supplementary documentation
- Quality of analysis and focus on key points and comparisons
- Clear, professional narrative

Assignment 3

Curation Initiative Proposal Presentation - 35 points, due June 7

Objectives:

- Construct a well-scoped and convincing presentation for a data curation or open data initiative.
- Gain experience designing curation advances for a designated community and goal.
- Demonstrate competency identifying and explaining data curation concepts and best practices for producing sharable and reusable data for a specific audience and purpose.

Scenario:
You are a data curator and have been invited by your institution to propose a new or expanded data service or initiative that would improve data access and use for a particular community or constituency. Your goal is to pitch a useful and feasible initiative or project. You will propose a plan that considers the particular needs and the context of the organization. The proposal will be practical, addressing factors involved in implementation, such as technical and expertise requirements, resources, and challenges. The project you propose may be inspired by or based on a model elsewhere. It could be an enhancement of content and/or functionality of an existing data resource, or it could be a new advancement that will create or improve data access and/or use. Your presentation should be designed for delivery to an audience that includes administrators who are curation novices, as well as professional peers who are in a good position to assess the merits and feasibility.

Focus

The proposal should be focused on collecting or enhancing data collections to meet a specified goal for a particular service community. You can be creative, but be sure to target a realistic community and goal.

For example, a community could be chemists in a corporate research unit, a linguistics department at a university, qualitative social scientists at a research institute, staff of an NGO, a local citizens group, a local community organization, or a professional association.

Some example goals, for illustrative purposes, might be:

- Recruiting and preparing data for to solve a community problem or support a research group

- Depositing all the data generated by a retiring researcher
- Systematic enrichment or cross-walking of metadata to improve discovery
- Improving documentation for high potential data to promote reuse
- Strategy for identifying and preparing “dark data” for deposit
- Aggregating distributed data in a specific problem area
- Constructing a data inventory and catalog for a complex organization

Guidelines:

Your proposal should draw on relevant topic areas and materials from class--data types and formats; metadata and documentation; identifiers; citation; data management plans; data services; policy and ethics; sustainability--where appropriate. But, you will also need to tailor your proposal to address the needs and interests of the specific target community.

The presentation of your proposed initiative should cover the following:

1. The problem the initiative will address, and the community that will benefit and how.
2. Specific data sources involved; their value to the community and specified purpose, characteristics, and challenges.
3. Stages of work: processes, tools, expertise, infrastructure, collaborations, etc.
4. Rough estimate of resources, costs, and timeline.
5. Sources – at least 6-8 recommended readings or supporting sources. Half should come from the course content, and the other half must be new materials related to your aims and user community.

Deliverable:
An 8-minute presentation of your proposal, with no more than 8 slides. The first slide should include the title, but also clearly indicate the problem and community addressed. The last slide should list the 6-8 of recommended resources. The presentation should be professional, suitable for sharing with potential employers as part of a portfolio.

Submit a file or a link to your presentation under ‘Assignments’ on Canvas. Please use the following file naming convention: [last name]_A[assignment #]_DC_Sp21.docx (e.g. Smith_A3_DC_Sp21....)

Grading rubric:

- Content: well selected, accurate, with high value to the community
- Narrative: well-paced, convincing, logical
- Design: clean visual presentation, precise language, helpful use of color, images, icons, etc.
- Recommended resources: authoritative, accessible, and appropriate for the audience

Lab Memos – 10 points

Two out of 4 Lab Memos required. Others, if completed, will be considered extra credit.
Memos should be informal but professional, about a paragraph or up to about 150 words.

Data Dictionaries, due April 7.

Based on the Data Dictionaries Lab Activity for Module 2:

Write a short memo on the purpose and value of data dictionaries and how they might differ in content and use across disciplines. Feel free to share any interesting or instructive examples.

Data Inventories, due April 21

Based on the Data Inventories Lab Activity for Module 4, write a Lab Memo considering these questions:

- What are the strengths of the New Orleans inventory approach. What could be enhanced or improved?
- What challenges would you expect to encounter carrying out such an inventory?

DMPTool, due May 3

Based on the Lab Activity for Module 6, write a short memo documenting your observations from reviewing the UM examples, the UW shared plans, and the DMP templates and tool functionality, more generally.

How self sufficient do you think researchers can be using the DMPTool?
What additional guidance or support, if any, would be most valuable for researchers in creating high quality data management plans?

Preservation collaboration, due May 19

Based on the Lab Activity for Module 8, on tools and solutions sharing in digital preservation community, write a short memo that responds to the following scenario and questions:

You are an information professional in a university library or government agency. In your role, you are responsible for the curation and preservation of data resources but dependent on the institutional IT unit for infrastructure and technology solutions.

- What limitations or misconceptions might IT personnel have regarding preservation?
- How would you collaborate with them to evolve a robust preservation system?

Concepts of the Week – 3 postings = 5 points total

For the following modules, select a concept from the readings for that week that you think is highly important or interesting. You may select concepts from required or recommended readings.

Post the concept and a short explanation of why you selected it on the Concepts Discussion Forum, by noon the day of class. Be prepared to talk briefly about your concept posting in class.

April 5 - Data Sharing and Access Across Disciplines

April 26 - Data Identity, Provenance, and Trust

May 17 - Preservation and Rescue

Participation – 15 points

Participation is an important part of the course, and it is essential to being a good citizen of the class.

Most participation for the class will take place during synchronous class time, through **interactions around lectures and discussion sessions, where we will take up other topics that arise through your lab activities and assignments**. Additional interaction will take place on the Discussion Forums. In particular, we will keep an active **Concepts of the Week** forum focused on the course readings. There are 3 required posting for particular dates that will result in assignment credit, but additional participation on this forum is strongly encouraged. It is a relatively light weight way to engage on core themes and ideas throughout the course.

The **Current Awareness** forum is for sharing general materials and resources on data curation and related topics. There are no expectations for participation on this forum, but your postings will contribute to your participation grade.

Overall, you should have an active voice in the classroom and on the forums. Being a good citizen requires coming to class, completing the readings in advance, and engaging constructively with the course material and respectfully with your colleagues. Responses to your colleagues' contributions, in class or on the discussion forums, must be respectful and constructive. You should aim to extend, clarify, or add other perspectives to ideas shared.

Participation grading: Your grade will be based on the quality and consistency of your contributions to class sessions and the discussion forums. You should strive to be part of the conversations as they unfold. "Outstanding" participation offers consistent, constructive contributions that integrate with, but do not dominate, class conversations.

University of Washington Information School

Data Curation II

Spring 2021

Schedule

EVENT	DATE	DESCRIPTION	COURSE MATERIAL
Module	03/29/2021 Monday	Week 1: Course Overview and Introduction [video] [content] [exercise]	Required Reading: ○ Course Content
Due	**03/30/2021 23:59 Tuesday**	**Sign Up for 15-minute Meet & Greet**	Where: ○ Canvas Calendar (see Course Overview Lecture) ○ (This isn't a hard deadline - just be aware that sign-ups are happening)
Due	**04/04/2021 23:59 Sunday**	**Introduce Yourself on Canvas**	Where: ○ Canvas Discussion Board
Due	**04/04/2021 23:59 Sunday**	**Read Assignment Overview**	
Module	04/05/2021 Monday	Week 2: Tables, Trees, & Triples [video] [content] [exercise]	Required Readings: ○ Course Content ○ Data's Shape

			○ Data on the Web - Best Practices ○ Read one additional Data on the Web - Use Case from this list Suggested Readings: ○ Abiteboul, S., Buneman, P., & Suciu, D. (2000). Data on the Web: from relations to semistructured data and XML
Due	**04/11/2021 23:59 Sunday**	**Week 2 Exercise Due**	Where: ○ Canvas Discussion Board
Due	**04/11/2021 23:59 Sunday**	**Post Data Pitch**	
Module	04/12/2021 Monday	Week 3: Tidy Data [video] [content] [exercise]	Required Readings: ○ Course Content ○ Rowson and Munoz (2016) Against Cleaning ○ Wickham, H. (2014), "Tidy Data," Journal of Statistical Software, 59, 1–23 Suggested Readings: ○ Wickham, H. (2014), "Tidy Data," Journal of Statistical Software (more code & examples than required reading) ○ Tierney, N. J., & Cook, D. H. (2018). Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations ○ Leek (2016) Non-tidy data ○ Broman, K. W., & Woo, K. H. (2018). Data organization in spreadsheets. The American Statistician, 72(1), 2-10 ○ Tort, F. (2010). Teaching spreadsheets: Curriculum design principles ○ Mack, K., Lee, J., Chang, K., Karahalios, K., & Parameswaran, A. (2018, April). Characterizing scalability issues in spreadsheet software using online forums ○ Formatting data tables in spreadsheets: Data Carpentry Lesson
Due	**04/18/2021 23:59 Sunday**	**Week 3 Exercise Due**	Where: ○ Canvas Discussion Board

Due	**04/18/2021 23:59 Sunday**	**Submit Statement of Work**	
Module	04/19/2021 Monday	Week 4: Data Integration [lecture] [guest lecture] [content] [exercise]	Required Readings: ○ Course Content Highly Recommended Readings: ○ Whong, Chris (2020) "Taming the MTA's Unruly Turnstile Data" ○ Wikipedia article on data integration Optional Readings: ○ Halevy, A., Rajaraman, A., & Ordille, J. (2006, September). Data integration: The teenage years. In Proceedings of the 32nd international conference on Very large data bases (pp. 9-16) ○ Abiteboul, S., Buneman, P., & Suciu, D. (2000). Data on the Web: from relations to semistructured data and XML. Morgan Kaufmann
Due	**04/25/2021 23:59 Sunday**	**Week 4 Exercise Due**	Where: ○ Canvas Discussion Board
Due	**04/25/2021 23:59 Sunday**	**Submit Users and Use Cases**	
Module	04/26/2021 Monday	Week 5: Data Packaging [video] [content] [exercise]	Required Readings: ○ Course Content ○ Bechhofer, S., De Roure, D., Gamble, M., Goble, C., & Buchan, I. (2010). Research objects: Towards exchange and reuse of digital knowledge ○ Skim this list of projects and tools for data packaging: Google Sheet or PDF ○ Neylon (2017) Packaging data Pick 1 of following to read or review in-depth: ○ BagIT for packaging QDR Data (social science / qualitative data) ○ Frictionless data (general and open data) ○ Data Crate (general research data) ○ Data Package (Ecology)

			○ HTRC Data Capsules (Digital Humanities)
Due	**05/02/2021 23:59 Sunday**	**Submit Collection Policies**	
Module	05/03/2021 Monday	Week 6: Repository Architectures [video] [content]	Required Readings: ○ Course Content ○ description of digital libraries (cyberinfrastructure) from the National Science Foundation program ○ This post from the IQSS staff at Harvard's Dataverse provides an excellent table comparing existing data repository services. Pay attention to the categories being compared, and how this related to the affordances of the software ○ Fallaw, C., Dunham, E., ... (2016). Overly honest data repository development. Code4Lib Review documentation for just one repository platform listed below (be sure to also look at an example of the platform's deployment): ○ Samavera (Open-source repository for universities and institutional repositories) ■ About ■ Technical Stack ■ Example deployment ○ Dataverse (Open-source repository for social science data) ■ About ■ Documentation ■ Example deployments https://data.qdr.syr.edu/ and https://dataverse.tdl.org/ ■ See the QDR Core Seal Trust documentation for more details on how Dataverse is configured ○ Fedora (Open-source repository with semantic capabilities - often used by science repositories) ■ About ■ Specifications ■ Developer Wiki ■ Example Deployment ■ ADS certification documentation for further info on how they use Fedora

- CKAN (open-source data repository - often used for civic data)
  - About
  - Documentation
  - Example deployments https://data.gov.au/ and Data.gov
  - Some additional info on Data.gov.au's CKAN
- Clowder (Open-source for long-tail data)
  - Description of project
  - Description of technical design
  - Example deployment

Suggested Readings:

- Amorim, R. C., Castro, J. A., Da Silva, J. R., & Ribeiro, C. (2017). A comparison of research data management platforms: architecture, flexible metadata and interoperability. Universal Access in the Information Society, 16(4), 851-862
- Lnenicka, M. (2015). An in-depth analysis of open data portals as an emerging public e-service. International Journal of Social, Education, Economics and Management Engineering, 9(2), 589-599. (see table 3 in particular for a comparative approach to Open Data portal evaluation)
- Cornell University Library Repository Principles and Strategies Handbook (I highly recommend this if you are looking for some background on how a University Library strategizes around digital infrastructures)
- Blanke, T., & Hedges, M. (2013). Scholarly primitives: Building institutional infrastructure for humanities e-Science. Future Generation Computer Systems, 29(2), 654-661

Module	05/10/2021 Monday	Week 7: Data Acquisition, Search, and Discovery [video] [content] [exercise]	Required Readings: ○ Course Content ○ Google Dataset Search: Building a search engine for datasets in an open Web ecosystem. ○ Facilitating the discovery of public datasets ○ Discovering millions of datasets on the web Suggested Readings: ○ Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories. ○ Understanding data search as a socio-technical practice. ○ Scientific user requirements for a herbarium data portal. ○ Scholar-built collections: A study of user requirements for

			(Humanities) research in large-scale digital libraries. ○ Improving the discoverability and web impact of open repositories: techniques and evaluation. Case Study (Optional): ○ A Data-Driven Approach to Appraisal and Selection at a Domain Data Repository.
Due	**05/16/2021 23:59 Sunday**	**Week 7 Exercise Due**	Where: ○ Canvas Discussion Board
Due	**05/16/2021 23:59 Sunday**	**Submit Transformations and Quality Assignment**	
Module	05/17/2021 Monday	Week 8: Metadata Application Profiles [video] [content] [exercise]	Required Readings: ○ Course Content ○ Application profiles: ■ Heery, R., & Patel, M. (2000). Application profiles: mixing and matching metadata schemas. Ariadne, (25) ■ The Singapore Framework for Application Profiles Note this is currently under revision by DCMI. You can catch up on their work here (and also see an example of use cases in the wild) ○ Some examples of metadata application profiles: ■ DPLA ■ Cornell Library ■ Carnegie Hall Archives Suggested Readings: ○ Hebron, T. K. (2018). Extending and Adapting Metadata Audit Tools for Mountain West Digital Library Members Code4Lib Journal, (41) ○ Curado Malta, M., Bermúdez Sabel, H., Baptista, A. A., & González-Blanco García, E. (2018). Validation of a metadata application profile domain model ○ Stein, A., & Dunham, E. (2018). Meaningful Data Sharing: Developing the Illinois Data Bank Metadata Framework. Journal of Library Metadata, 18(2), 59-83

Due	**05/23/2021 23:59 Sunday**	**Submit Metadata Application Profile**	
Module	05/24/2021 Monday	Week 9: Linked Data [video] [content]	Required Readings: ○ Course Content ○ Allemang, D., & Hendler, J. (2011). Semantic web for the working ontologist: effective modeling in RDFS and OWL. Second Edition ■ Read Chapter 1 for an introduction to SW's concepts. If you are interested Chapter 2 gives a bit more detail on how the SW works, and Chapter 3 introduces RDF and knowledge modeling. ○ Ontology Development 101 (Noy and McGuiness) ■ Read Section 1 and 2; (3 and 4 are optional) ■ Note - this is a classic formulation of what an ontology is and how to create one. The software they reference in building out the example is called Protege (free https://protege.stanford.edu/). If you are really keen you can follow along. (For reference - this short list from Wikipedia is quite helpful.) ○ Ontology for Data Science ○ Semantic Web for the Legal Domain Suggested Readings: ○ ARL White Paper on Wikidata: Opportunities and Recommendations (2019)
Due	**05/30/2021 23:59 Sunday**	**Submit Licensing Assignment**	
Module	05/31/2021 Monday	Week 10: Emerging Topics [content]	Required Reading: ○ Course Content
Due	**06/03/2021 23:59 Thursday**	**Optional Repository Assignment**	

Due	06/06/2021 23:59 Sunday	Submit Final Protocol

Box 352840 - Mary Gates Hall, Ste. 370
Seattle
Washington

uw_ischool
ischool.uw.edu/

LIS 598 | **Digital Preservation** | Winter 2020

Reading List + Syllabus

Week 1

Required

- Trevor Owens (2018) Theory and Craft of Digital Preservation
    - Chapter 1: Beyond Digital Hype and Digital Anxiety
- ITHAKA S+R State of Digital Preservation 2018 report https://sr.ithaka.org/publications/the-state-of-digital-preservation-in-2018/ (note trends that seem unique to current environment, and those that seem longstanding)

Case Study

- Arroyo-Ramirez, Elvia (2016) Invisible Defaults and Perceived Limitations: Processing the Juan Gelman Files.

Background reading (optional):

- Ross, S. (2012). Digital preservation, archival science and methodological foundations for digital libraries. New Review of Information Networking, 17(1), 43-68. http://www.tandfonline.com/doi/pdf/10.1080/13614576.2012.679446
- Tibbo, H. R. (2003). On the nature and importance of archiving in the digital age. Advances in Computers, 57, 1-67. [PDF]
- Lavoie, Brian, and Lorcan Dempsey. “Thirteen Ways of Looking At... Digital Preservation.” D-Lib Magazine 10, no. 7/8 (July/August 2004). http://dx.doi.org/10.1045/july2004-lavoie.
- Conway, P.. (2010). Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas. *The Library Quarterly: Information, Community, Policy*, *80*(1), 61–79. http://doi.org/10.1086/648463Thibodeau, K. (2002). Overview of technological approaches to digital preservation and challenges in coming years. In The state of digital preservation: An international perspective (pp. 4-31). Washington, DC: Council on Library and Information Resources. Retrieved from http://www.clir.org/pubs/reports/pub107/pub107.pdf
- Waters, D., & Garrett, J. (1996). Preserving Digital Information. Report of the Task Force on Archiving of Digital Information. The Commission on Preservation and Access, 1400 16th St., NW, Suite 740, Washington, DC 20036-2217. http://www.clir.org/pubs/reports/pub63

- Conway (1996) Preservation in the Digital World. CLIR Report. http://www.clir.org/pubs/reports/conway2/index.html/

If you're interested in an International context to digital preservation this is a nice point of comparison for the North American focused work above:

- European Commission 2020 report (statement on digital preservation responsibilities of member states): http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:283:0039:0045:EN:PDF
- http://ec.europa.eu/digital-agenda/en/news/2015-national-reports-digital-preservation

Week 2

Required

- Trevor Owens (2018)
- Theory and Craft of Digital Preservation
    - Chapter 3: Understanding Digital Objects.
    - Chapter 7: Managing Copies and Formats.
- NDSA (2014) Checking Your Digital Content: An NDSA Report on Fixity Best Practices https://ndsa.org/documents/NDSA-Fixity-Guidance-Report-final100214.pdf

Case Study

- Morfin, Jo Ana The Changing Landscape: Preserving Mexican Digital Documentary Heritage. iPres 2018 https://osf.io/teyk9/

Optional (but highly recommended)

- Brown, A. (2008). "Selecting File Formats for Long-Term Preservation ".
- Pearson, D. &Webb, C. (2007). Defining File Format Obsolescence: A Risky Journey . IJDC 3 (1) p. 89-106

Optional (but recommended)

- Trevor Owens (2018) Theory and Craft of Digital Preservation
    - Chapter 2: Artifact, Information, or Folklore: Preservation's Divergent Lineages
- Moore, R. (2008). Towards a theory of digital preservation . *International Journal of Digital Curation*, *3*(1), 63-75. Pay attention to what he says are the "steps" of file-level characterization.

After you do the readings from this week browse content at the following links:

- Library of Congress - Sustainability of Digital Formats Planning for Library of Congress Collections
- Pronom - Technical Registry of the National Archives (UK)

Week 3

Required

- Phillips, M., Bailey, J., Goethals, A., & Owens, T. (2013). The NDSA Levels of Digital Preservation: An Explanation and Uses . IS&T Archiving, Washington, USA.
- _Lavoie, B, (2014) The Open Archival Information System Reference Model: Introductory
  Guide. DPC Technology Watch Series Report,(read Section 5, pages 7-20)
  https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file

Case Study

- Archiving a Revolution in the Digital Age, Archiving as an Act of Resistance
  https://www.ibraaz.org/essays/163/

Optional Readings (all below)

- Trevor Owens (2018) Theory and Craft of Digital Preservation
  - Chapter 5: The Craft of Digital Preservation

Applications of OAIS (optional)

- McDonough, J. P. (2012, January). 'Knee-Deep in the Data': Practical Problems in Applying the OAIS Reference Model to the Preservation of Computer Games. In System Science (HICSS), 2012 45th Hawaii International Conference on (pp. 1625-1634). IEEE. PDF
- Vardigan, M., & Whiteman, C. (2007). ICPSR meets OAIS: applying the OAIS reference model to the social science archive context. Archival Science, 7(1), 73-87. PDF

Digital Curation Centre Lifecycle

- Higgins, S. (2008). The DCC curation lifecycle model. International Journal of Digital Curation, 3(1), 134-140. PDF
- Ball, A. (2009). Preservation and curation in institutional repositories. PDF (Note this is a much larger report - just review the DCC lifecycle section unless otherwise inclined.)

Variable Media Questionnaire

- Ippolito, Jon. “Accommodating the Unpredictable: The Variable Media Questionnaire.” In Permanence Through Change: The Variable Media Approach, edited by Alain Depocas, Jon Ippolito, and Caitlin Jones, 47-53. New York and Montréal: The Solomon R. Guggenheim Museum and Daniel Langlois Foundation, 2003. PDF
- Sterling, Bruce. “Digital Decay.” In Permanence Through Change: The Variable Media Approach, edited by Alain Depocas, Jon Ippolito, and Caitlin Jones, 11-22. New York and Montréal: The Solomon R. Guggenheim Museum and Daniel Langlois Foundation, 2003. PDF

Storage Calculator Links

- General File Size Calculator http://jan.ucc.nau.edu/lrm22/pixels2bytes/calculator.htm
- Video Space Calculator https://www.digitalrebellion.com/webapps/videocalc
- Audio Storage Calculator http://www.theaudioarchive.com/TAA_Resources_File_Size.htm
- Digitization equations http://preservationtutorial.library.cornell.edu/technical/technicalC-05.html

Week 4

Required

- Trevor Owens (2018) Theory and Craft of Digital Preservation
- Chapter 8: Arranging and Describing Digital ObjectsCaplan, P. (2009, December). Understanding PREMIS. Washington DC, USA: Library of Congress. https://www.loc.gov/standards/premis/understanding-premis.pdf (Read sections 1-3)
- Dappert, A., Guenther, R. S., & Peyrard, S. (2016). Digital Preservation Metadata for Practitioners. Implementing PREMIS. Cham, Switzerland: Springer. Chapter 2: How to Develop a Digital Preservation Metadata Profile: Risk and Requirements Analysis (PDF)

Case Study

- Hasegan, T. C. (2018) The opportunities and challenges in managing indigenous digital archives: The case of the Indigenous Subject Guide at the CNAIR https://rising.globalvoices.org/blog/2018/12/14/the-opportunities-and-challenges-in-managing-indigenous-digital-archives/

Optional - Practical Case Studies

- DeRidder, J. L. (2015) Generating Standardized Audio Technical Metadata: AES57. Journal of Code4LIB. http://journal.code4lib.org/articles/10828

- Wilson, A. (2010). How Much Is Enough: Metadata for Preserving Digital Data . *Journal of Library Metadata*, *10*(2-3), 205-217.

Sites to Browse

- PREMIS (PREservation Metadata: Implementation Strategies).

Standards to Review

- METS - Metadata Encoding & Transmission Standard
- MODS - Metadata Object Description Standard
- PREMIS Data Dictionary for Preservation Metadata

Week 5

Required

- Sheldon, M. (2013). Analysis of Current Digital Preservation Policies: Archives, Libraries and Museums. The Library of Congress. http://www.digitalpreservation.gov/documents/Analysis%20of%20Current%20Digital%20Preservation%20Policies.pdf
- Becker, C., Kulovits, H., Guttenbrunner, M., Strodl, S., Rauber, A., & Hofman, H. (2009). Systematic planning for digital preservation: evaluating potential strategies and building preservation plans. International journal on digital libraries, 10(4), 133-157. PDF
- Bishoff, L. (2010). Digital Preservation Plan. Ensuring Long Term Access and Authenticity of Digital Collections. Information Standards Quarterly 22 (2). **PDF**
    - Note that the Bishoff reading is a short framework document that we will use as the basis for our Digital Preservation plans this quarter - so if you read anything this week make sure it is that!

Example Digital Preservation Plans (skim based on your interest)

Higher Ed

- Yale University Digital Preservation Plan

- Dartmouth

- UMass Amherst

Data Repositories & Digital Libraries

- ICPSR

- HathiTrust

- British National Library

_-

Odum Institute

Case Study

Case- Preservation Planning for Emerging Formats at the British Library https://osf.io/p92u8/

Optional

- Rafferty, E., & Pad, B. (2017). Better Together: A Holistic Approach to Creating a Digital Preservation Policy in an Art Museum _. *Art Documentation: Journal of the Art Libraries Society of North America*, *36*(1), 149-162.

Week 6

Required

- Audit and certification (Digital Preservation Handbook): https://www.dpconline.org/handbook/institutional-strategies/audit-and-certification
- Read background on certification of digital repositories: https://www.coretrustseal.org/why-certification/
- Skim the Core Seal Trust (CST) Application from QDR (note I wrote this) https://www.coretrustseal.org/wp-content/uploads/2018/11/Qualitative-Data-Repository.pdf _ __ In class we will discuss the application. Be prepared to describe: Continuity of access, confidentiality, licensing etc

Choose one of the following to read:

- Business Cases, Costs, and Benefits for Preservation: https://www.dpconline.org/handbook/institutional-strategies/business-cases-benefits-costs-and-impact

- 20 Questions about the cost of digital preservation: https://www.metaarchive.org/public/publishing/ma_20costquestions_final.pdf?thumblink

- Checklist for writing a business case for digital preservation: http://wiki.dpconline.org/index.php?title=IOE_case_study:_Check_list_for_writing_a_business_case_for_digital_preservation

Case Study

- Are we there yet? Understanding digital preservation costs and benefits https://www.dpconline.org/blog/are-we-there-yet

Optional

- Vermaaten, S., Lavoie, B., & Caplan, P. (2012). Identifying threats to successful digital preservation: The SPOT model for risk assessment. *D-Lib Magazine*, *18*(9), 4. http://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html
- Phillips, M., Bailey, J., Goethals, A., & Owens, T. (2013). The NDSA Levels of Digital Preservation: An Explanation and Uses. *IS&T Archiving, Washington, USA*. http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_Levels_Archiving_2013.pdf
- Literature Review of Digital Preservation Models (from organizational standpoint) : http://www.jeffersonbailey.com/i-review-6-digital-preservation-models-so-you-dont-have-to/

For Reference:

- Research Data Alliance (RDA) Working Group on Repository Certification
- Catalog of Core Requirements (for data repositories): https://docs.google.com/document/d/1_DPwSA5P8LpK9Q34BhxJmX8So2GKL7eSLa-G-z5JvVg/edit
- Procedures for basic certification https://docs.google.com/document/d/1KvcPc8siwLI_e8nCMYI5Jhp955ybjWZEMNB39XPO11U/edit

Optional (case studies)

- Houghton, B. (2015). Trustworthiness: Self-assessment of an Institutional Repository against ISO 16363-2012. *D-Lib Magazine*, *21*(3), 5. http://www.dlib.org/dlib/march15/houghton/03houghton.html
- Ravenwood, C., Muir, A., & Matthews, G. (2015). Stakeholders in the Selection of Digital Material for Preservation: Relationships, Responsibilities, and Influence. *Collection Management*, *40*(2), 83-110.
- Lacinak, C. (2015) What to Do After Digitization. *ARSC Guide to Audio Preservation*, 127. http://www.clir.org/pubs/reports/pub164/pub164.pdf#page=139

Week 7

Read for class:

- The Librarians Saving the Internet**:** **https://apps.sciencefriday.com/data/librarians.html**

- More Archive, Less Process - A Blog Post from the Library of Congress on creating federal web archives https://blogs.loc.gov/thesignal/2018/08/more-web-archives-less-process/

- Costa, M., Gomes, D., & Silva, M. J. (2017). The evolution of web archiving. International Journal on Digital Libraries, 18(3), 191-205. https://link.springer.com/article/10.1007/s00799-016-0171-9 (Read this entire paper if you are looking to invest your time and attention in web archiving- Strategically read if not)

PDF

Examples of GUI-based Web Archiving:

- http://rhizome.org/

- https://cobwebarchive.org/

Case Study:

- Davis, C. (2014). Archiving the Web: A Case Study from the University of Victoria. *code {4} lib Journal* *https://journal.code4lib.org/articles/10015*

Optional (but recommended):

- Web Archiving Environmental Scan https://dash.harvard.edu/bitstream/handle/1/25658314/print_HL_web_archiving_env_scan_2017.pdf?sequence=4

- Leetaru, K. (2019) Preserving Online News In An Ephemeral Web: A Look At Four Months Of Global Digital Journalism. Forbes. https://www.forbes.com/sites/kalevleetaru/2019/01/05/preserving-online-news-in-an-ephemeral-web-a-look-at-four-months-of-global-digital-journalism/#7698ce8f2ce8

Week 8

Read for class:

- Acker, A., & Kriesberg, A. (2017). Tweets may be archived: Civic engagement, digital preservation and Obama white house social media data. Proceedings of the Association for Information Science and Technology, 54(1), 1-9.

PDF

- NARA Best Practices
https://www.archives.gov/files/records-mgmt/resources/socialmediacapture.pdf

- See the Documenting the Now resource (we will go over in our demo) https://www.docnow.io/

See this tweet (rant) from a digital preservationist at the Internet Archive about the precarity of using YouTube as an archive: https://twitter.com/textfiles/status/1219418143300386823 (and a good example of how this plays out in the real world : https://twitter.com/E_I_T/status/1219426295831879680

Case Study:

- How We Collected WeChat
https://www.vam.ac.uk/blog/international-initiatives/how-we-collected-wechat

Optional (but recommended):

- (Very short but cool example of using AI to enhance a video)
https://www.digg.com/2020/arrival-train-la-ciotat-upscaled

- Lischer-Katz, Z. (2017 Jan.). Studying the materiality of media archives in the age of digitization: Forensics, infrastructures, and ecologies. First Monday, 22(1).

http://firstmonday.org/ojs/index.php/fm/article/view/7263

- Drake J. (2016) Expanding #ArchivesForBlackLives to Traditional Archival Repositories
https://medium.com/on-archivy/expanding-archivesforblacklives-to-traditional-archival-repositories-b88641e2daf6

- Littman, J, (2018) Islamic State Extremists Are Using the Internet Archive to Deliver Propaganda
https://medium.com/on-archivy/islamic-state-extremists-are-using-the-internet-archive-to-deliver-propaganda-a132597dd16

- Morris, J (2018) Pod Save Podcasts: Saving and Studying Podcasting's Booming Sonic Culture https://bellocollective.com/pod-save-podcasts-ca5bb2fb1cb3

Week 9

Required
- O'Meara, E. (2017). Personal Digital Archiving: DPC Technology Watch Report 15-01. *The American Archivist*, *80*(1), 240-243. Read Sections 3, 4, and 5 - https://www.dpconline.org/docs/technology-watch-reports/1460-twr15-01/file

- "Silicon Valley Can't Be Trusted With Our History: We create almost everything on the internet, but we control almost none of it." https://www.buzzfeednews.com/article/evanhill/silicon-valley-cant-be-trusted-with-our-history

- St-Esprit, M (2018) How Millennial Parents Are Reinventing the Cherished Family Photo Album. The Atlantic. https://www.theatlantic.com/family/archive/2018/12/preserving-kids-photos-in-the-digital-age/577579/

Case Study:

- Documenting Ferguson - an interview with Meredith Evans: http://www.processhistory.org/documenting-ferguson/

- See the full site and pay attention to the way that material can be submitted to the archive: http://digital.wustl.edu/ferguson/

Optional (but recommended):

- Reply All Podcast: Episode 71 "The Picture Taker" https://www.gimletmedia.com/reply-all/71-the-picture-taker

- Elsden, C., Kirk, D. S., & Durrant, A. C. (2016). A quantified past: Toward design for remembering with personal informatics. *Human–Computer Interaction*, *31*(6), 518-557. https://www.tandfonline.com/doi/full/10.1080/07370024.2015.1093422

- Jääskeläinen, A., & Uosukainen, L. (2018). Citizen Archive: My Precious Information. *New Review of Information Networking*, 1-11. PDF

- Hester, J.L. (2018) The Quest for a Universal Translator for Old, Obsolete Computer Files. Atlas Obscura https://www.atlasobscura.com/articles/how-to-open-old-computer-files

Week 10

- Mtima, L. (2015). Copyright and social justice in the digital information society: three steps toward intellectual property social justice. *Hous. L. Rev.*, *53*, 459.

https://heinonline.org/HOL/Page?collection=journals&handle=hein.journals/hulr53&id=481&men_tab=srchresults

Case Study:

- Orland, Kyle (2018) Game industry pushes back against efforts to restore gameplay servers https://arstechnica.com/gaming/2018/02/preservation-or-theft-historians-publishers-argue-over-dead-game-servers/
    - Also see -> No legal way to preserve MMOs http://motherboard.vice.com/read/museums-can-now-legally-jailbreak-game-consoles-but-gamers-arent-allowed
    - And this older article on some early efforts https://www.engadget.com/2014/01/04/the-game-archaeologist-four-efforts-to-preserve-dead-mmos/

Suggested readings

Challenges due to IP rights in new media archiving:

- TPP vs. Municipal Archives http://www.geekman.ca/2015/11/the-tpp-vs-municipal-archives.html
- Mike Masnick. Happy Birthday And The Problem With The Copyright Office's 'Orphan Works' Plan. https: //www.techdirt.com/articles/ 20151005/18115432443/happy-birthday-problem-with-copyright- offices-orphan-works-plan.shtml
- US gov't grants limited right to revive games behind "abandoned" servers : http://arstechnica.com/gaming/2015/11/u-s-govt-grants-limited-right-to-revive-games-behind-abandoned-servers/
    - A new update: http://www.ipwatchdog.com/2018/01/27/dmca-exemption-online-video-game-preservation/id=92913/

Visual Archives (IP Challenges)

- Besser, H. (2016). 25 years of digital archives of visual materials: what we've done, what we've learned, and what challenges remain. *Visual Studies*, *31*(2), 95-108. https://www.tandfonline.com/doi/abs/10.1080/1472586X.2016.1173888

DCMA

- Brewster Kahle. Internet Archive Helps Secure Exemption To The Digital Millennium Copy- right Act. https://archive.org/post/ 82097/internet-archive-helps-secure-exemption-to-the-digital- millennium-copyright-act

- International Rights Statements for Cultural Heritage Digital Objects: http://dp.la/info/2015/10/06/whitepapers-for-establishing-international-and-interoperable-rights-statements-released/

Orphans, Copyright and Legal Rulings on IP in Digital Preservation

- Copyright in Mass Digitization http://copyright.gov/orphan/reports/orphan-works2015.pdf
- DRM and Libraries http://www.ala.org/advocacy/copyright/digitalrights
    - Copyright and libraries in 2018 http://www.ip-watch.org/2018/02/16/copyright-libraries-around-world-2018/
- Legal and Policy Aspects of Software Preservation https://saaers.wordpress.com/2016/09/26/software-preservation-network-legal-and-policy-aspects-of-software-preservation/
- Suthersanen, U. (2018). Who owns the orphans? Property in digital cultural heritage assets. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3139933

Intro to API for open government data

An Yan & Bree Norlander

Description

The ability to collect data is essential for a data curator. This lab session will introduce students to the concept of an API and provide examples of and practical guidelines for accessing data and metadata from public sources using APIs. Specifically, we will first learn to work with SODA API[1] provided by an Open Government Data (OGD) Portal provider, Socrata. We will learn how to retrieve open government datasets and their metadata through the SODA API, Discovery API, and Metadata API.

Learning Objectives

- Understand the concept of an API
- Use an API to collect data from public data sources such as open government data portals and open research data repositories.

Prerequisites

- Basic knowledge of any programming language would be helpful, but not required.
- Basic understanding of open data and data repositories; knowledge of popular open data portal software providers (CKAN, Socrata, Dataverse, etc.)
- Familiarity with common web exchange file formats, such as JSON, XML, and CSV.

Lab notes

- Need laptop and internet access

Introduction to API

API stands for Application Programming Interface. An API defines how different applications or different components of software talk to each other, providing a way for developers to simplify programming. Just as a graphical user interface makes it easier for people to use programs, APIs make it easier for developers to use certain functions or data from another application (or another component of an application) without knowing the implementation details [1][2].

In this lab, we will focus on web APIs. Some applications such as Facebook, Twitter, and Google Map provide services and data through Web APIs. The API is the part of the application that receives requests and sends responses. For example, if you are an iPhone user and you use Yelp to search for a restaurant, you will see an Apple Map embedded in Yelp showing you the location of the restaurant. Yelp does not provide map services itself, but it

[1] https://dev.socrata.com/consumers/getting-started.html

uses Apple map's API to integrate with its own functions. Another example of API usage is the Google or Facebook log-in capability on other platforms. When you sign up for a non-Google, non-Facebook website, they may ask if you want to sign in with a Facebook or Google account. This website is using Facebook or Google API to retrieve your account information.

The APIs that will be used in this lab adhere to an architectural style called REST (Representational State Transfer). "REST services tend to offer an easy-to-parse URL structure consisting of nouns that reflect the logical hierarchical categories of the data on offer" (Maddox, 2014). An example of a REST API web URL is: http://api.us.socrata.com/api/catalog/v1/domains. When using a REST API, you will need to know how to structure your query. The rules for doing so can usually be found in the API documentation of a web application.

Why should we learn about API? Software developers use APIs to create well-structured applications. Data scientists and data curators use APIs to collect large amounts of data automatically. Learning to access data and metadata using an API is the main objective of this lab.

Example Overview

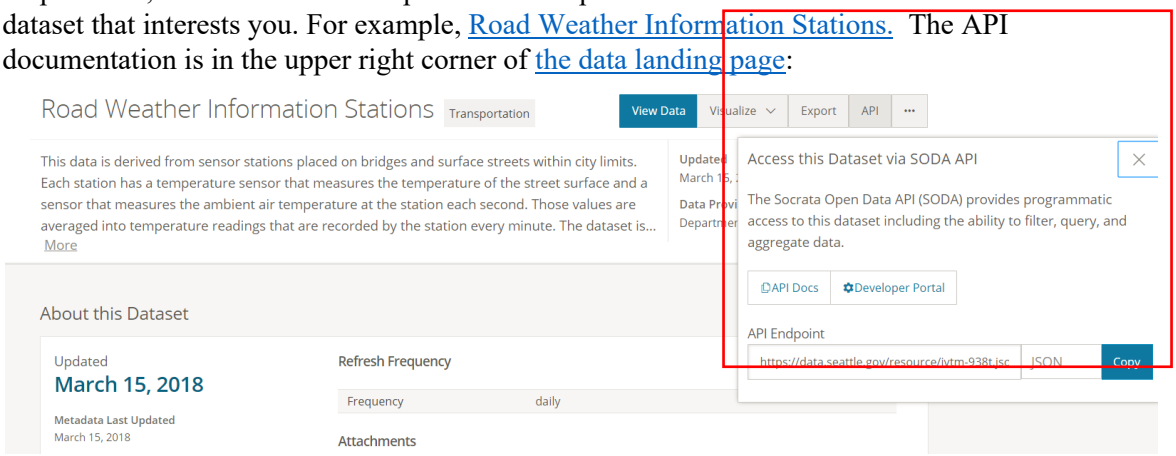
I will walk through two examples of retrieving datasets from the municipal open data portal of City of Seattle using an API:

- Example 1: Retrieve open data from Open Government Data Platforms using an API
- Example 2: Retrieve Open Government Data metadata using API

The underlying software, Socrata, provides APIs for accessing, filtering, and downloading data and metadata.

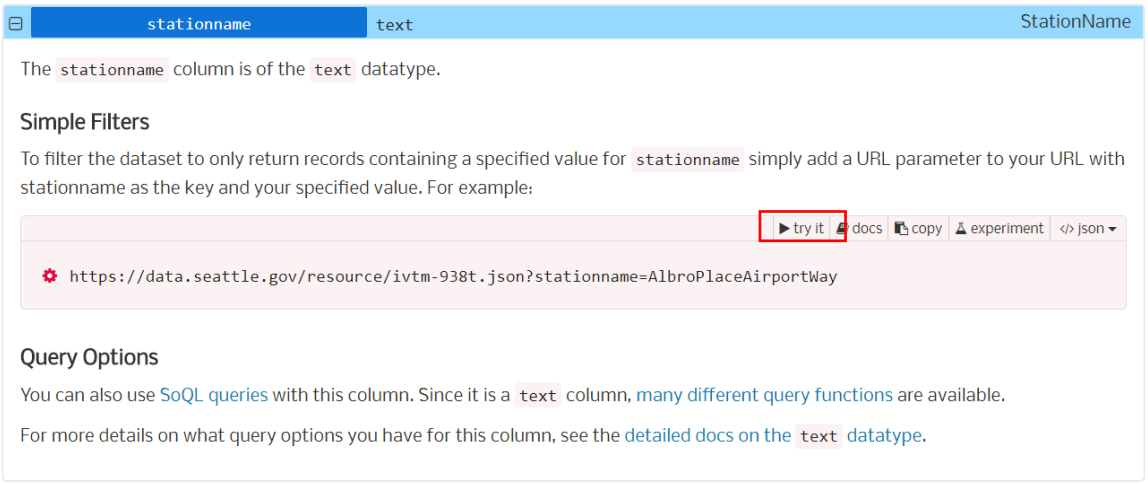
Example 1: Retrieve data from Socrata Platforms

- Step 1: First, take some time to explore Seattle Open Data Portal datasets and choose a dataset that interests you. For example, Road Weather Information Stations. The API documentation is in the upper right corner of the data landing page:



- Click "API" and copy the "API Endpoint" using the default JSON data format to a new tab or window in your. Try this again for a different format of the data such as CSV. Finally, click on "API" again, but this time choose the "API Docs" link.

- Step 2: on the “API Docs” page, read the “Fields” section and try different filters. For example, you can filter the data by specifying the parameter “stationname”. Click “try it” to see what response (data) you get.

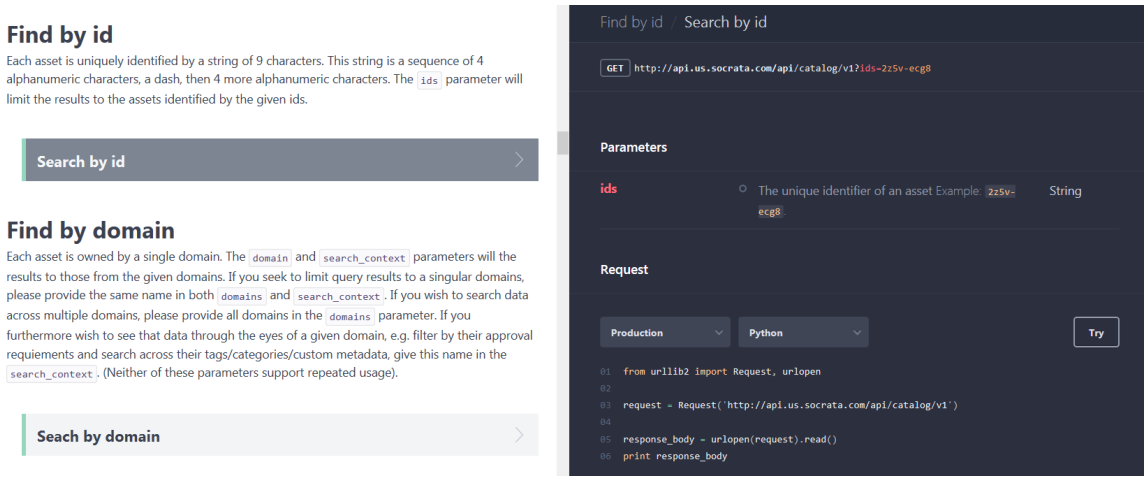


- Step 3: Go to https://dev.socrata.com/docs/endpoints.html, review API endpoint concept, and try the examples on the page.

Example 2: Retrieve catalog information and metadata using Socrata’s API.

Example 2a: Explore catalog-level metadata using Discovery API

- Step 1: Go to ‘Discovery API, read the introduction, and scroll through the rest of the content. Try a few examples by clicking on the text boxes that begin with “Search by . . .”. You can also view different code examples by choosing from the drop-down menu under “Request” that is set by default to “Raw”. Take a look at a few code examples, especially if you are familiar with any of the provided languages.



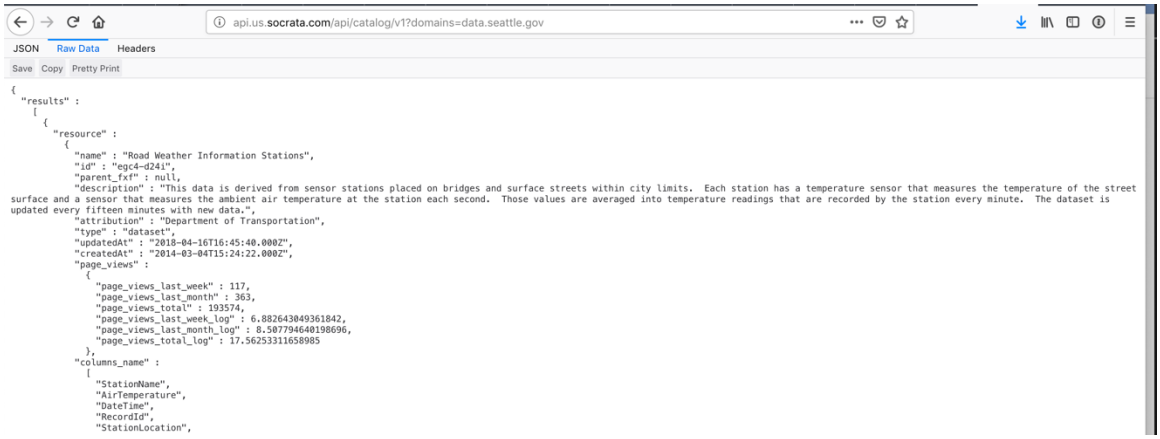
- Step 2: After trying some of the examples in the documentation, try customizing your own API call in a browser window. Here are some examples:

1) Find the list of domains and the resources count for each Socrata-hosted domain using the following API call (you can copy and paste this into a browser window): http://api.us.socrata.com/api/catalog/v1/domains.


```
{
  "results" :
    [
      { "thing" : "2014bonds.cityofws.org", "count" : 5 },
      { "thing" : "amopen.amo.on.ca", "count" : 78 },
      { "thing" : "bchi.bigcitieshealth.org", "count" : 85 },
      { "thing" : "bea.data.commerce.gov", "count" : 3 },
      { "thing" : "bis.data.commerce.gov", "count" : 2 },
      { "thing" : "brigades.opendatanetwork.com", "count" : 507 },
      { "thing" : "bronx.lehman.cuny.edu", "count" : 773 },
      { "thing" : "bythenumbers.sco.ca.gov", "count" : 103 },
```


"Thing" represents portals hosted by Socrata and "count" represents the number of resources each portal possesses.

You can then search within a particular domain. Here is an example: http://api.us.socrata.com/api/catalog/v1?domains=data.seattle.gov (results using Firefox shown below).



Try it yourself with a different domain.

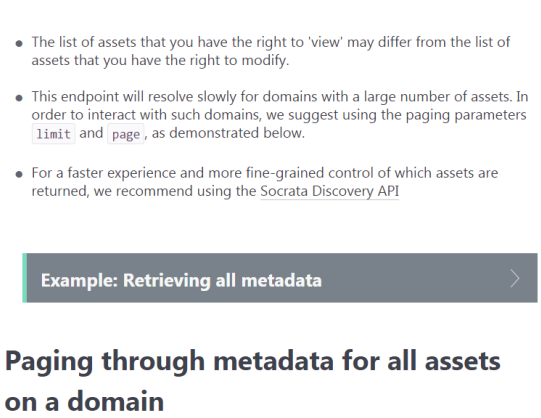
2) Get all metadata of resources that belong to category "public safety" using the following API call:http://api.us.socrata.com/api/catalog/v1?categories=public%20safety . As we can see, the results contain metadata records that belong to category "public safety" of all Socrata-hosted portals.

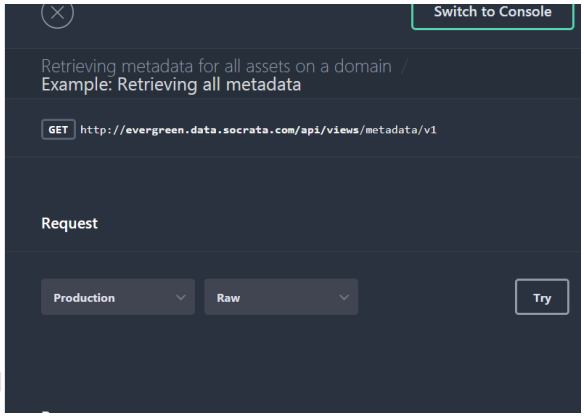
```
{
  "results" :
    [
      {
        "resource" :
          {
            "name" : "2014 Traffic Volumes Overview Map",
            "id" : "9kha-7xbc",
            "parent_fxf" : [ "2efd-hq2e", "a5rb-dcy9" ],
            "description" : "Traffic Volumes (also known as Traffic Counts) ar
Districts Offices.  The Districts calculated the volumes hourly, daily, and mc
westbound and \"Ahead\" describes northbound and eastbound.",
            "attribution" : "California Department of Transportation",
            "type" : "map",
            "updatedAt" : "2016-08-27T19:20:25.000Z",
            "createdAt" : "2016-03-07T17:31:17.000Z",
            "page_views" :
              {
                "page_views_last_week" : 0,
                "page_views_last_month" : 0,
                "page_views_total" : 288,
                "page_views_last_week_log" : 0.0,
                "page_views_last_month_log" : 0.0,
                "page_views_total_log" : 8.17492568250068
              },
```


Try the search again with a different category. You can find a list of all possible categories with this API call: http://api.us.socrata.com/api/catalog/v1/categories.

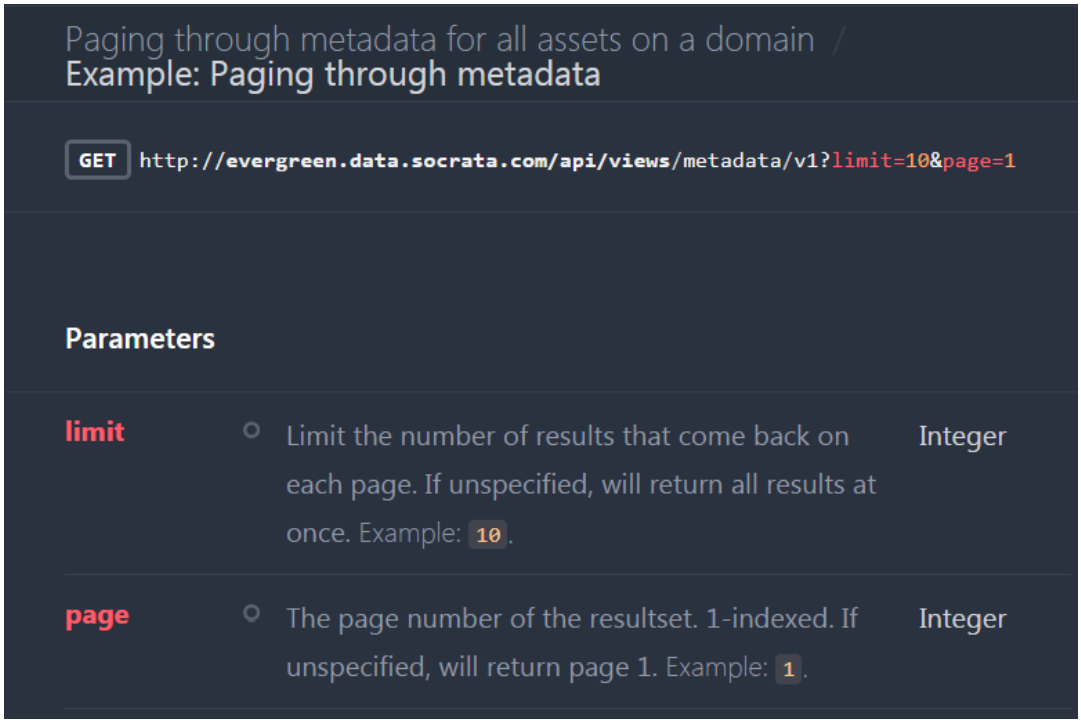
Example 2b: Retrieve asset-level metadata of datasets using Metadata API

- Step 1: Go to Socrata's Metadata API page. Read the introduction and try a few examples on this page. For example, you could choose to retrieve all metadata on a domain (see image below).





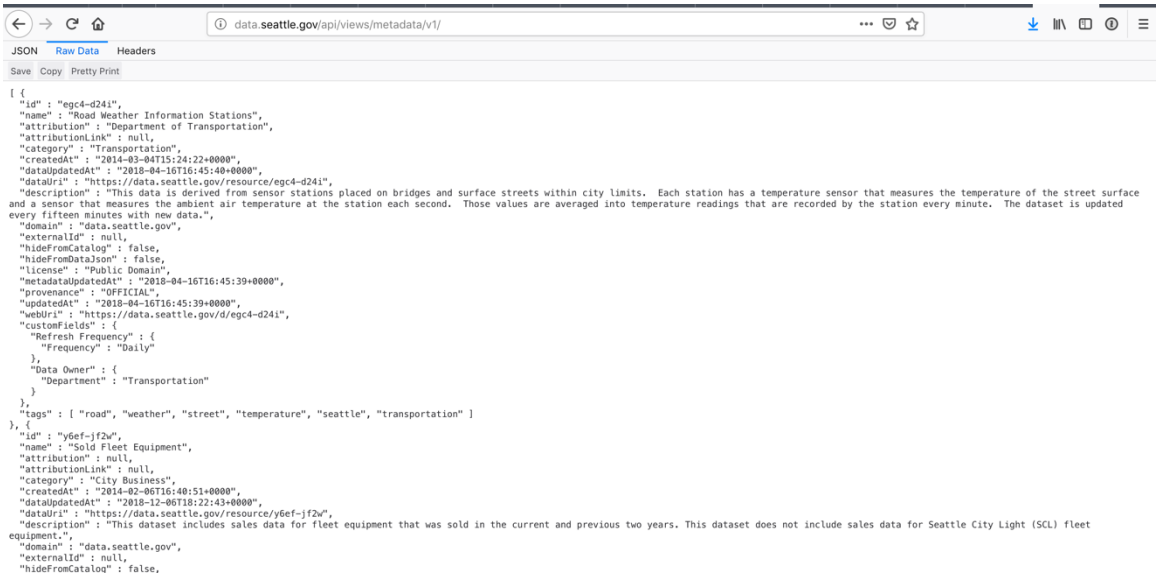
- Step 2: Customize your API calls to retrieve metadata of interest to you. Add different parameters. For example, you can start by working with existing examples provided by Socrata.



The API call, http://evergreen.data.socrata.com/api/views/metadata/v1?limit=10&page=1 will return the 10 metadata records of page one from https://evergreen.data.socrata.com.


```
[ {
  "id" : "b5gk-7v6a",
  "name" : "America in Crisis: Opioid Addiction",
  "attribution" : null,
  "attributionLink" : null,
  "category" : "Public Health",
  "createdAt" : "2016-03-21T17:47:12+0000",
  "dataUpdatedAt" : "2016-03-21T17:47:12+0000",
  "dataUri" : "https://evergreen.data.socrata.com/resource/b5gk-7v6a",
  "description" : "Learn about the national, state, county, and city perspectiv
what cities and counties across the country are doing to address the issue.",
  "domain" : "evergreen.data.socrata.com",
  "externalId" : null,
  "hideFromCatalog" : false,
  "hideFromDataJson" : false,
  "license" : null,
  "metadataUpdatedAt" : "2017-06-26T16:19:27+0000",
  "provenance" : "OFFICIAL",
  "updatedAt" : "2017-06-26T16:19:27+0000",
  "webUri" : "https://evergreen.data.socrata.com/stories/s/b5gk-7v6a",
  "customFields" : {
    "Publisher" : {
      "Agency Name" : "Public Health"
    }
  },
  "tags" : null
```

Try these queries with a different domain that you discovered in Example 2a Step 1. For example, we can see the metadata from the seattle domain with the following API call: http://data.seattle.gov/api/views/metadata/v1/ (results using Firefox shown below). Notice how we replaced evergreen.data.socrata.com in the URL with data.seattle.gov.




```
[ {
  "id" : "egc4-d24i",
  "name" : "Road Weather Information Stations",
  "attribution" : "Department of Transportation",
  "attributionLink" : null,
  "category" : "Transportation",
  "createdAt" : "2014-03-04T15:24:22+0000",
  "dataUpdatedAt" : "2018-04-16T16:45:40+0000",
  "dataUri" : "https://data.seattle.gov/resource/egc4-d24i",
  "description" : "This data is derived from sensor stations placed on bridges and surface streets within city limits.  Each station has a temperature sensor that measures the temperature of the street surface
and a sensor that measures the ambient air temperature at the station each second.  Those values are averaged into temperature readings that are recorded by the station every minute.  The dataset is updated
every fifteen minutes with new data.",
  "domain" : "data.seattle.gov",
  "externalId" : null,
  "hideFromCatalog" : false,
  "hideFromDataJson" : false,
  "license" : "Public Domain",
  "metadataUpdatedAt" : "2018-04-16T16:45:39+0000",
  "provenance" : "OFFICIAL",
  "updatedAt" : "2018-04-16T16:45:39+0000",
  "webUri" : "https://data.seattle.gov/d/egc4-d24i",
  "customFields" : {
    "Refresh Frequency" : {
      "Frequency" : "Daily"
    },
    "Data Owner" : {
      "Department" : "Transportation"
    }
  },
  "tags" : [ "road", "weather", "street", "temperature", "seattle", "transportation" ]
}, {
  "id" : "y6ef-jf2w",
  "name" : "Sold Fleet Equipment",
  "attribution" : null,
  "attributionLink" : null,
  "category" : "City Business",
  "createdAt" : "2014-02-06T16:40:51+0000",
  "dataUpdatedAt" : "2018-12-06T18:22:43+0000",
  "dataUri" : "https://data.seattle.gov/resource/y6ef-jf2w",
  "description" : "This dataset includes sales data for fleet equipment that was sold in the current and previous two years. This dataset does not include sales data for Seattle City Light (SCL) fleet
equipment.",
  "domain" : "data.seattle.gov",
  "externalId" : null,
  "hideFromCatalog" : false,
```


Resources:

1. REST API: https://en.wikipedia.org/wiki/Representational_state_transfer
2. Other useful APIs:
   a) Twitter APIs.
   b) PubMed API
   c) Dryad API
   d) CKAN API
3. Data Science project using Soda API: https://github.com/ViDA-NYU/urban-data-study

Exercises:

1. Explore Soda API and discovery API documents.
2. Try to use Soda API to access datasets of interest, filter data records according to your selected criteria. Submit your API calls.

   e.g. Get all records about "Animal Complaints" after April 4th, 2018:
   https://data.seattle.gov/resource/pu5n-trf4.json?$where=event_clearance_date >'2018-04-04T12:00:00.000'&initial_type_group=ANIMAL COMPLAINTS

3. Try to use Soda discovery API or metadata API to get catalog information or asset-level metadata information about open government data portals. Use filters and parameters to refine your queries. Submit your API calls.

   e.g. *http://api.us.socrata.com/api/catalog/v1?categories=public%20safety*

   e.g. http://data.seattle.gov/api/views/metadata/v1?limit=10&page=1

Creator:

This material is part of Open Data Literacy Project funded by IMLS grant.

An Yan: Information School, University of Washington

Bree Norlander: Information School, University of Washington

Carole Palmer: Information School, University of Washington

References and Credits:

1. https://www.mulesoft.com/cn/resources/api/what-is-an-api
2. https://en.wikipedia.org/wiki/Application_programming_interface#Web_APIs
3. Geary, David M. *Core JavaServer Faces, 3/e*. Pearson Education India, 2004.
4. Maddox, S. (2014, February 16). API Types. Retrieved February 19, 2018 from https://ffeathers.wordpress.com/2014/02/16/api-types/

Working with civic data portal metadata with OpenRefine

An Yan 05/24/2018

Description

As a data curator, we constantly need to deal with messy data and metadata. OpenRefine is a popular open-source tool for cleaning and transforming data. This tutorial will show you how to use OpenRefine to clean metadata about Open Civic Portals pulled from Socrata, an Open Government Data (OGD) Portal provider. We will first use Discovery API provided by Socrata to get metadata. Then we will learn to use OpenRefine to perform format transformation, duplicates checking, columns splitting, and exploratory analysis.

Learning Objectives

- Use API to collect data from public data sources such as open government data portals.
- Use OpenRefine to do data cleaning.

Prerequisites

- Basic understanding of open data and data repositories; knowledge of popular open data portal software providers (CKAN, Socrata, Dataverse, etc,.)
- Familiarity with common web exchange file formats, such as JSON, XML, and CSV.
- Basic knowledge of web API.

Lab notes

- Need Laptop and internet access

Introduction to OpenRefine

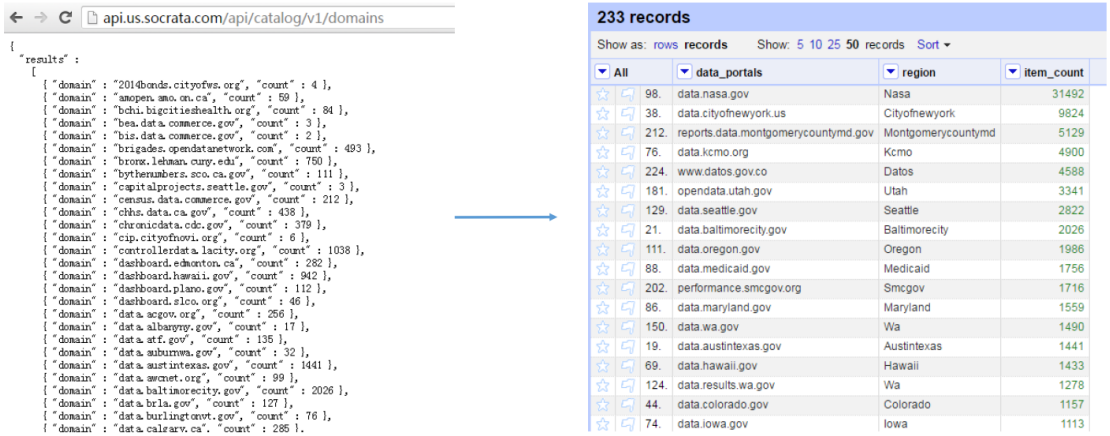
OpenRefine is similar Google spreadsheet in some sense. They both handle tabular data in a browser and allow sharing among different users. However, OpenRefine can deal with many data formats that are frequently used on web environment, such as HTML, JSON, XML. Those are the ones that excel spreadsheet cannot easily deal with.

We can follow the download link in this page to get OpenRefine. Once we have it installed, click the icon and our browser will open automatically showing the OpenRefine workspace.

Example

This example is about maintaining a table of all open portals (most of them being open government portals) supported by Socrata, the number of resources each portal hosts, and the

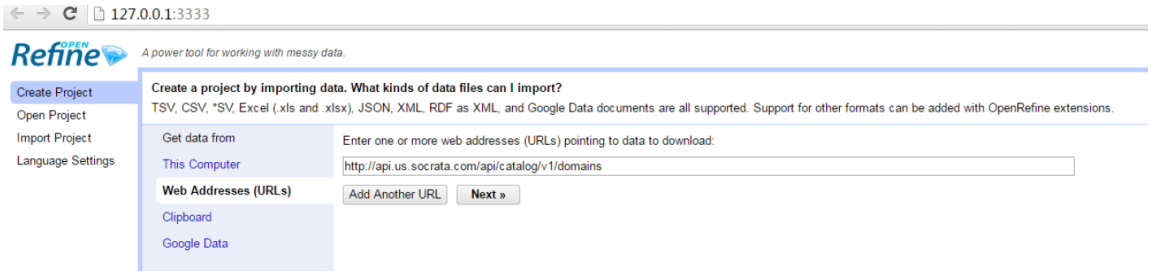
regional government or organization each portal represents. The data used in this example can be retrieved from Socarata API in JSON format using http://api.us.socrata.com/api/catalog/v1/domains. We will use OpenRefine to transform it into a nice tabular format.



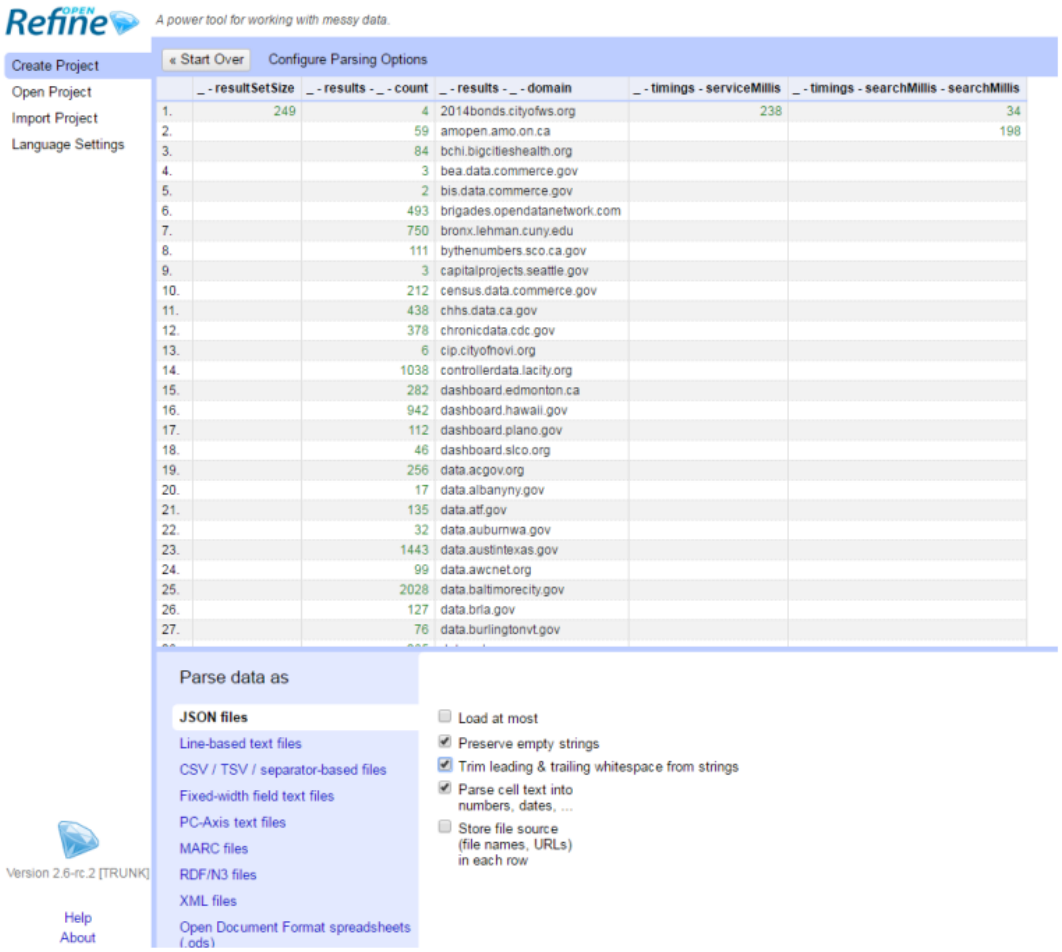
Please refer to our previous tutorial on "Introduction to API for Open Government Data" for more details on data retrieval.

Step 1: Create Project

In your browser, choose create project from "Web Addresses (URLs)" and paste the API call http://api.us.socrata.com/api/catalog/v1/domains . This allows OpenRefine to retrieve data directly from API so that we do not need to download the data.

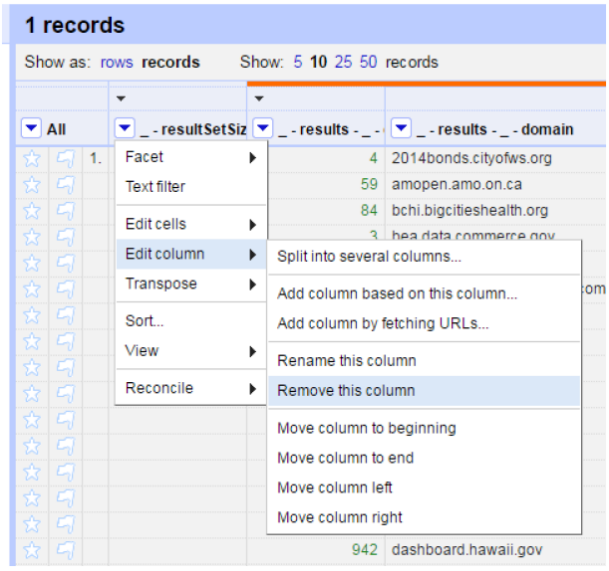


Check the following boxes: "Parse cell text into numbers, dates" and "Trim leading & trailing whitespace from strings" and click "Create Project".



Step 2: Remove and rename columns

Now we can see that the JSON format data has been converted into a table. But there are columns from the original data that we do not need, for example, the “resultSetSize” element. We can delete that column by clicking the small triangle at the column name, choose “Edit column” -> “Remove this column”.



Using “Edit column” -> “Rename this column”, we could clean up column names. In this example, columns are renamed to “data_portals” and “Item_count” from “results_domain” and “result_count”, respectively.

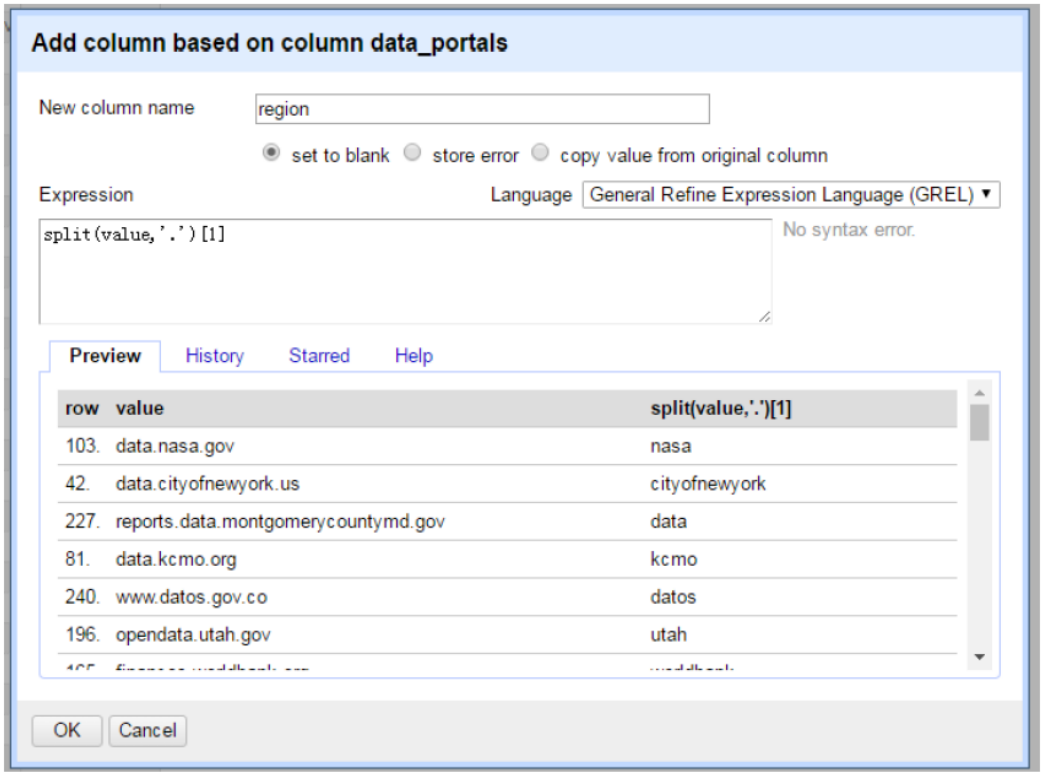
Step 3: Check duplicates

OpenRefine makes it very easy to check duplicates based on columns. Use “Facet” -> “Customized facets” -> “Duplicates facet”, we could see in the left-hand panel, there is not any duplicate in “data_portals” column.

Step 4: Add columns and split string

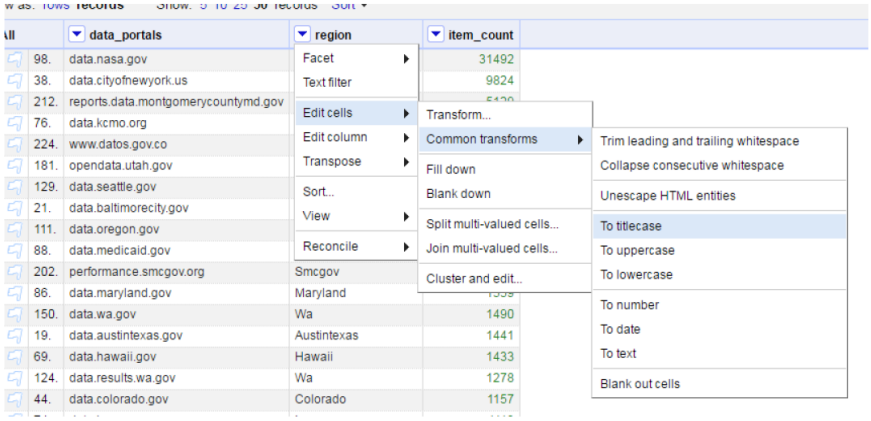
Now we want to add another column called “region” to store the names of the government or the organization hosting each portal, for example, NASA and Seattle. We could work from the urls and extract region names from them. For example, we could extract “seattle” from “data.seattle.gov”. The way we could achieve this is by splitting the urls by “.” (resulting in three elements: “data”, “seattle”, “gov”), and take the second element (“seattle”) out. We can use “Add

column based on another column" to create a new column "region" based on "data_portals" column. To do that, we use the "General Refine Expression Language (GREL)" to implement splitting. Type in "split(value, '.')[1]" in "Expression" box to pull out the second element of the url and assign it to column "region". For more on GREL, please visit the OpenRefine wiki page.



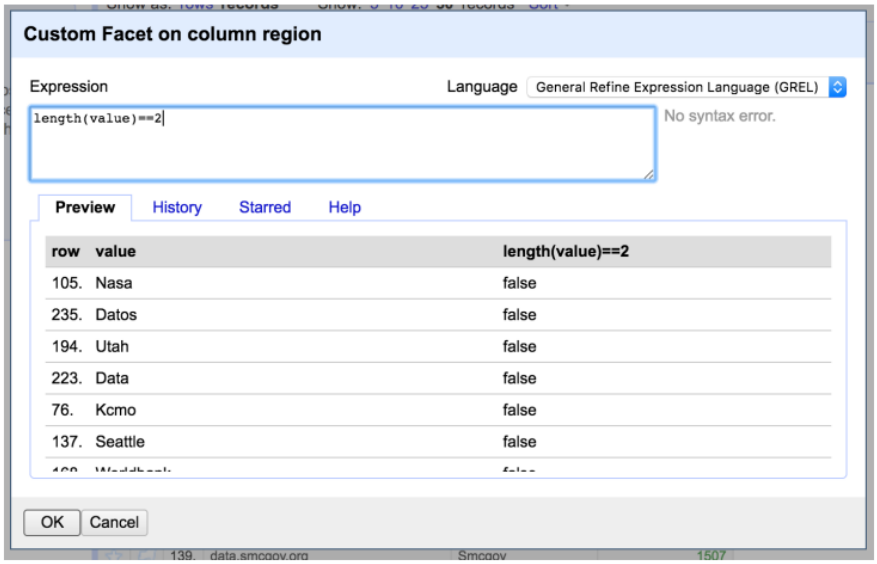
Step 5: Common transforms

After extracting region names out from urls, we may want to format them, for example, transform them into title case. OpenRefine allows for easy transformation by providing a set of functionalities under "Edit cell" -> "Common transforms". Let us make all region names title case.

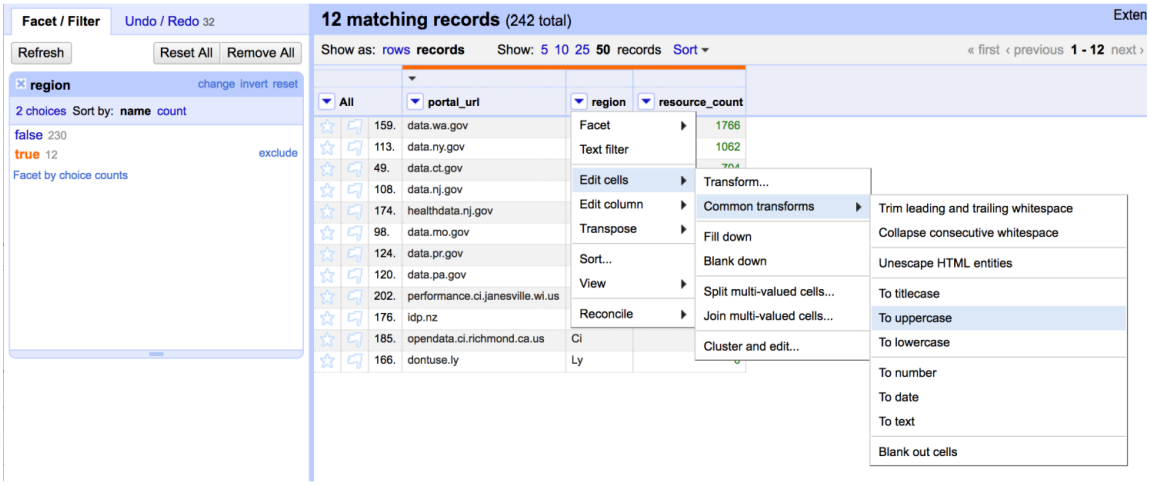


Step 6: Customized facets and transforms

We may also want to transform state abbreviations to upper case, for example, “Wa” to “WA”. To do this, we need to first filter out region names with length of two. Again, using “Facet” tool, we can create a “Custom text facet” to choose region names with two letters. The expression goes in the “Expression box” is “length(value) == 2”.

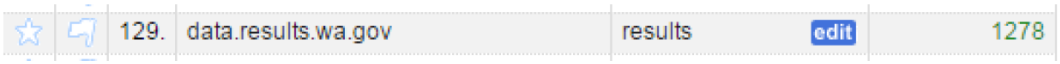


Then on the left-hand side panel, we will see that there are 12 matching records, we then use “Common transforms” to convert them to uppercase.



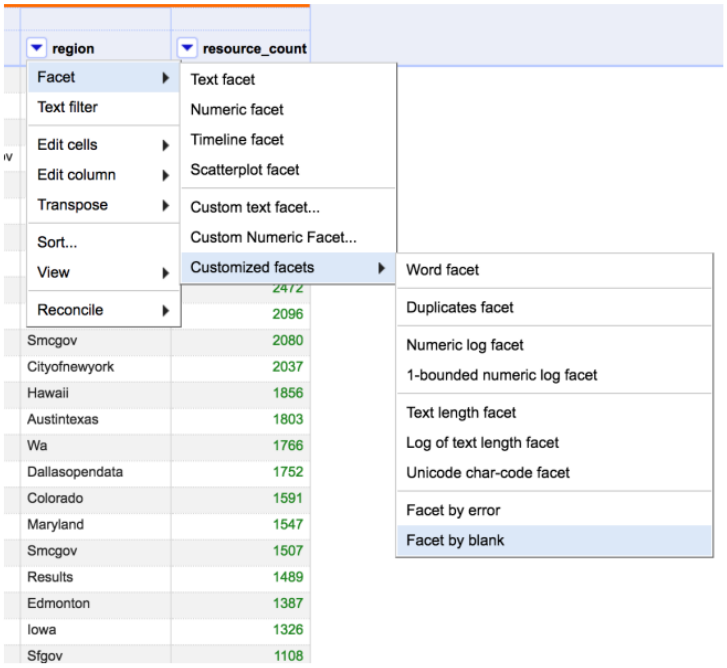
Step 7: Manually edit cell value

OpenRefine’s automatic data cleaning and transform functionalities have been very useful so far. However, there are still places that need manual editing. We can edit individual cell values by moving the cursor to the cell to be edited and click “edit”. We do this until every cell is satisfactory.



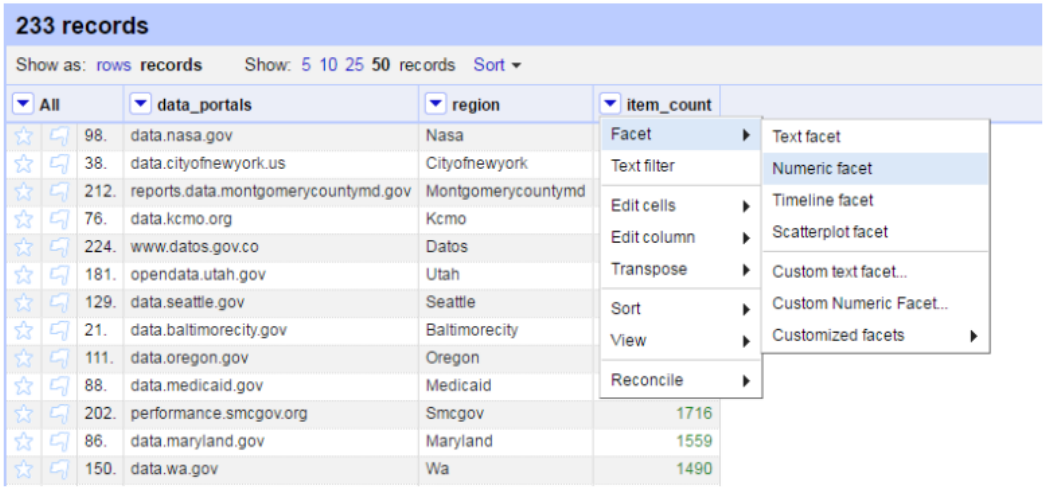
Step 8: Delete rows with blank cells

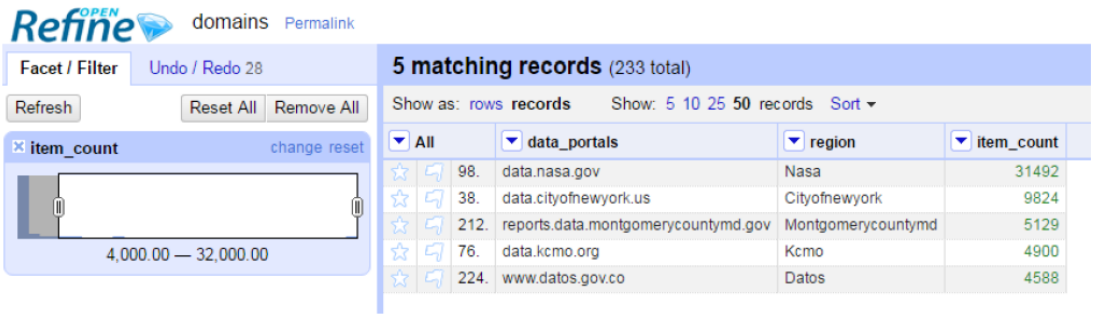
Lastly, we may want to remove the “bad values” or the “outliers”. In the original data, some urls are of invalid format. We could either manually inspect the entire dataset and remove them, or detect them using OpenRefine functionalities. For example, when we did step 4, we ran “split(value, ’.’)[1]” to extract the second element of urls. This step resulted in some blank cells since the urls are invalid, such as “stubox”. Therefore, they do not have a second element. By using the “Facet by blank” tool, we can detect those invalid urls.



Step 9: Exploratory analysis

We could use OpenRefine to do some basic exploratory analysis. For example, if we are interested in examining data portals with more than 4000 resources, we could first sort the table by “item_count”, and then use “numeric facet” to filter the records.





Step 10: Export and share

Finally, we have transformed a JSON format messy data into a nice table. We can export the resulting table into a variety of formats including Excel and use the “Permalink” in the left-upper part to share the workspace with others.

Resources

- OpenRefine Github: https://github.com/OpenRefine/OpenRefine
- OpenRefine Tutorial: http://enipedia.tudelft.nl/wiki/OpenRefine_Tutorial
- More on string operation and regex: https://github.com/OpenRefine/OpenRefine/wiki/Understanding-Regular-Expressions

Creators

This material is part of Open Data Literacy Project funded by IMLS grant.

An Yan: Information School, University of Washington

Bree Norlander: Information School, University of Washington

Carole Palmer: Information School, University of Washington

References and Credits:

1. http://openrefine.org/
2. https://socratadiscovery.docs.apiary.io/

Appendix 3:

Selected posters

Expanding Open Data Expertise in the Public Sector

Carole L. Palmer, Nicholas Weber, Bree Norlander, Kaitlin Throgmorton

ODL
OPEN DATA LITERACY

Public sector institutions are becoming active in open data.

Working with ODL has increased their ability to launch and sustain new initiatives.

INTRODUCTION

Open Data Literacy (ODL) improves accessibility and use of open data by advancing data literacy and expanding open data expertise in the public sector. To do this, ODL has developed iSchool graduate curriculum, coordinated open data fieldwork, produced open educational resources, analyzed library data published on city open data portals, and surveyed public libraries across the state on current activities, ambitions, and challenges in open data.

EXPANDED iSCHOOL CURRICULUM

- New courses and content on open data for the public sector
- Guest lectures by expert practitioners
- Hands-on learning with open data tools and technologies, such as GitHub, OpenRefine, and APIs

CURRENT OPEN DATA ACTIVITY IN WA

Collaborating with the Washington State Library, ODL surveyed 51 public libraries about open data activities and interests. Overall, WA libraries, large and small, are active or interested in promoting and working with open data.

- More than 60% agree open data aligns with their mission and community's interests
- 43% are active in open data reference assistance
- 33% are active in open data publishing

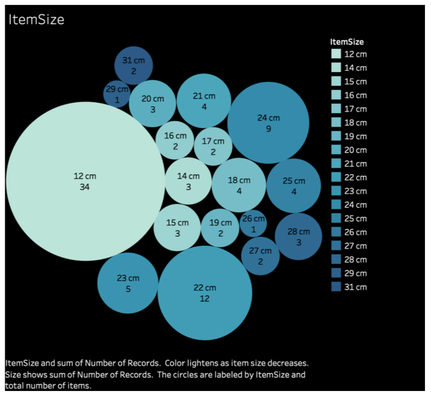
"Open data is an area where skilled guides are still needed. With the appropriate training, our librarians could be showing our community a richer way to view the world than through a simple google search."

— Senior librarian from a small county library (service population: 5,000 - 24,999)

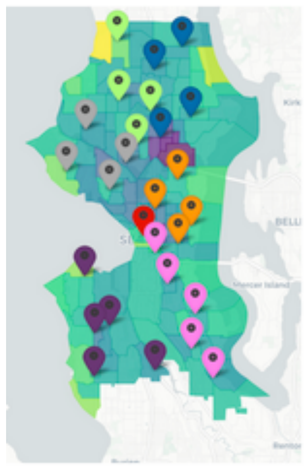
EXAMPLE FIELDWORK & OUTCOMES

Internship Projects

- Collected, cleaned, and published open data
- Analyzed library data



- Conducted analysis and assessment of open data catalogs, metadata standards, data management plans, public records requests, and library role in open data
- Developed dashboards for open data visualization
- Created open data engagement programs, such as conferences, instructional sessions, and guidebooks



Notable Student Accomplishments

- Employment offers from sponsors
- Co-authorship of scholarly publications
- Conference presentations at national conferences

EXPERTISE REALIZED

Key data curation competencies, as outlined by the National Research Council (2015), and open data expertise recognized in feedback from sponsors and interns.

Sponsor Acknowledgement of Intern Capability

- *Data Collection & Management*
  - Strong skills in research data management, scholarly data management, and government data management
- *Data Analytics*
  - Knowledge of coding
- *Services & Support*
  - Interpersonal and relational skills including communication, curiosity, enthusiasm, confidence, and self-directedness
- *Management & Administration*
  - Clear benefits in the cross-fertilization of academia and practice

Intern Feedback on Expertise Gained

- *Disciplinary & Institutional Data Practices*
  - Library principles and skills; open data environment and vocabulary; metadata experience; qualitative interviewing and focus groups; presenting and instructing on findings
- *Data Collection & Management*
  - Relational databases, repository creation
- *Data Analytics*
  - Data visualization, statistical analysis, unstructured data analysis
- *Technology, Tools, & Infrastructure*
  - R, Tableau, Zotero, Drupal, Excel, Python, Jupyter Notebooks, Github, SQL, OpenRefine, R Studio, Census API, Socrata API, PowerBI
- *Values & Principles*
  - Data ethics, FAIR principles, data privacy and security
- *Management & Administration*
  - Solutions for complex information problems

National Research Council. (2015). Preparing the Workforce for Digital Curation. Final Report of the Committee on Future Career Opportunities and Educational Requirements for Digital Curation, Board on Research Data. National Academies Press. 106 pp.

INSTITUTE of Museum and Library SERVICES

Funded by IMLS Laura Bush 21st Century Librarians Program.
Grant number: 67-5285.

Public Sector Partners:

SOS Office of the Secretary of State Washington State Library

Washington State Department of Transportation

The Seattle Public Library

ASOTIN COUNTY LIBRARY

UW Partners:

TASCHA

Advancing Open Data:

Aligning Education with Public Sector Data Challenges

Carole L. Palmer, Nicholas Weber, Bree Norlander, Kaitlin Throgmorton

clpalmer@uw.edu, nmweber@uw.edu, norlab@uw.edu, kthrog@uw.edu

Introduction

LIS education has invested heavily in workforce development for data services in research institutions, however little has been done on the open data competencies needed in public libraries and government agencies.

The Open Data Literacy (ODL) project is working to prepare students to build infrastructure, as well as curate and sustain data resources for the public. To do this, ODL advances data curation and data science curriculum, and provides field experiences in collaboration with Washington state and city agencies.

We are examining the current data problems faced by these agencies to help determine the expertise required to advance public sector open data initiatives and inform improvements in our curriculum.

Prominent Data Challenges

Primary data challenges, based on the first two years of student field experiences included:

- aggregating dispersed, inconsistent data
- improving user access and ability to view and interpret data and public records
- assessing and enhancing metadata, catalogs, indexing
- designing and delivering community and professional outreach

Focus Groups

Two post-internship focus groups surfaced important points about student preparation and scoping internship projects.

Areas for further coverage in curriculum:

- metadata
- storage and cloud services
- APIs
- legal and ethical dimensions
- public sector and not-for-profit organizations
- working openly with Github and other tools

Unanticipated effort in field experiences:

- learning local context, navigating government organizations
- framing productive stages of work
- communication and coordination with stakeholders
- local technology constraints – tools, support, coordination

Organizations and Projects

Requirements / Analysis Projects

WASHINGTON STATE DEPARTMENT OF TRANSPORTATION, summer internships

Project 1: ***Business Analysis of WSDOT Data or Term Search (DOTS)***

- Data catalog analysis to improve access and use by technical and business stakeholders.

Project 2: ***Analysis of Public Disclosure Requests***

- Open data recommendations based on analysis of frequently requested datasets and records.

CITY OF SEATTLE, summer internship

Project**:** ***Context for Communities: Metadata Standards for Civic Data***

- Assessment of civic metadata standards to identify best option for City of Seattle purposes.

WASHINGTON STATE OFFICE OF THE CHIEF INFORMATION OFFICER & CITY OF SEATTLE, summer internship

Project**:** ***Analysis of Public Records Requests***

- Analysis of public records requests to support proactive disclosure of datasets.

Development Projects

WASHINGTON STATE OFFICE OF THE CHIEF INFORMATION OFFICER, capstone; summer internship

Project 1: ***WA State Public Meetings Calendar***

- Established data standard and backend for visualization and user queries.

Project 2: ***OCIO Geospatial Program***

- Consolidated 3 geoportals through indexing and provenance tracking.

PUGET SOUND PARTNERSHIP (State of WA), capstone

Project: ***Opening Up the Data: Visualizing the Effectiveness of Puget Sound Restoration Efforts***

- Built processing pipeline and visualization for dispersed, interoperable restoration data.

WASHINGTON DEPARTMENT OF FISH & WILDLIFE, capstone

Project: ***Total Resource Observation Utility Team (TROUT)***

- Developed dashboard to visualize high-resolution public data to track Puget Sound land change.

WASHINGTON STATE DEPARTMENT OF ECOLOGY, capstone

Project: ***Waste 2 Resources***

- Transferred data to open portal and developed visualization tool for waste management.

Engagement Projects

SEATTLE PUBLIC LIBRARY, summer internship

Project: ***Open Data Unconference Framework***

- Curriculum on accessing, understanding, analyzing open data for public and library professionals.

WASHINGTON STATE LIBRARY, summer internship

Project: ***Open Data in Public Libraries***

- Analysis of local government open data and public libraries' roles in publishing and instruction.

CITY OF SEATTLE AND WASHINGTON STATE OFFICE OF THE CHIEF INFORMATION OFFICER, summer internship

Project: ***Open Data Alliance***

- Created guidebook for coalition of open data professionals concerned with public records requests.

Knowledge & Skills

(National Research Council, 2015*)

Disciplinary & Institutional Data Practices

- standards of evidence, quality, uncertainty
- data types and formats
- methods and instruments

Data Collection & Management

- sampling and acquisition techniques
- processing, transformation, documentation
- standards, metadata, ontologies, interoperability
- databases

Data Analytics

- statistics
- programming
- data mining
- visualization

Archiving & Preservation

- integrity
- transfer and migration
- preservation metadata

Technologies, Tools, & Infrastructure

- repository infrastructure
- network architecture
- web services
- system administration
- usability

Values & Principles

- open access
- access and use restrictions
- privacy

Services & Support

- liaison and consulting
- enhancement: metadata, annotation, linking
- instruction, training, outreach

Management & Administration

- policy and planning
- sustainability
- cost-benefit analysis
- cross-institutional coordination

*National Research Council. (2015). Preparing the Workforce for Digital Curation. Final Report of the Committee on Future Career Opportunities and Educational Requirements for Digital Curation, Board on Research Data. National Academies Press. 106 pp.

Curriculum Implications

Data curation and data science coursework offered by the UW iSchool covers many key areas of expertise. Reinforcement is needed in all areas related to access, use, and sustainability within Data Collection & Management, Data Analytics, and Archiving & Preservation. Further investment is needed in Technology, Tools, & Infrastructure, Values & Principles, Services & Support, and Management & Administration for students to succeed in open data positions in public sector organizations.

Acknowledgements: ODL is funded by a grant from the Institute of Museum and Library Services, Laura Bush 21st Century Librarians Program. Grant number: 67-5285.
The team also wishes to acknowledge An Yan for her important contributions to the project.

Public Sector Organizations

To find out more about ODL:
Please email odluw@uw.edu
Or visit our website: http://odl.ischool.uw.edu

Appendix 4:

Selected other outputs

Welcome to the Public Libraries and Open Data Survey

Welcome

Welcome to the Public Libraries and Open Data Survey.

This survey is being administered by Open Data Literacy (ODL), a project at the Information School at the University of Washington. Our aim is to learn about current interests and activities in open data among libraries across the state of Washington. To learn more about this survey and about open data in general, please visit https://odl.ischool.uw.edu/survey.

Thank you for taking this survey. Please click the blue arrow below to continue.

Survey Consent

Consent

Before beginning the survey, please read the terms outlined below. To participate, indicate your informed consent by responding to the following statements.

By checking this box, I confirm that I am 18 years of age or older.

By checking this box, I confirm the following:

- I have read and understand the purpose of the survey, and I am choosing to participate.
- I understand that my participation is voluntary.
- I understand that I can withdraw from the study at any time prior to submission of the survey, with no explanation needed.
- I understand that unless I choose to self-identify at the end of the survey, no identifiable information about me or my library will be recorded.
- I understand that deidentified survey data will be deposited in an open access repository.
- I understand that cookies may be placed on my browser by Qualtrics, the survey software, to enable various features.

The following statement(s) are about the level of anonymity to be maintained with your survey responses in reporting results.

I agree that excerpts from my open-ended responses may be used for reports, publications, and other research outputs.

Please select all that apply regarding the use of your excerpts:

(Note, if you choose to associate your name or institution with your responses, please provide the optional contact information at the end of the survey.)

My institution may be included in association with my excerpts.

My name may be included in association with my excerpts.

My excerpts may only be used anonymously.

If you have any questions, please contact Dr. Carole Palmer (clpalmer@uw.edu) or Kaitlin Throgmorton (kthrog@uw.edu) at the

University of Washington.

Please click the blue arrow to continue.

Instructions & Institutional Profile

Instructions

The survey will take approximately 15-20 minutes to complete. No prior background in open data is required to take the survey. If, however, you would like to know more about open data at any point before, during, or after the survey, we encourage you to visit the survey webpage and the Open Data Literacy website.

Please select the size of your library's service population:

Under 5,000

5,000 - 24,999

25,000 - 99,999

100,000 - 249,999

Over 250,000

Are you representing a library system?

Yes

No

If you'd like to clarify any details about your service population or organizational structure, please do so below:

Current Context

Current Context

Open data activities relate to many established library roles, including reference service, provision of digital collections, and community programming and engagement. Please indicate the degree to which you agree with the following statements.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Our library serves as a gateway to local government information.	○	○	○	○	○
Our library provides information on community	○	○	○	○	○

community activities and resources.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Our library provides access to local history or other special collections.	○	○	○	○	○
Much of our local and government information is available in digital form.	○	○	○	○	○
Our library offers guidance on how to find or use new information resources.	○	○	○	○	○
Our library offers guidance on the use of information technologies.	○	○	○	○	○
Community members turn to the library with questions about matters of local public interest.	○	○	○	○	○
Our local government turns to the library for help with projects	○	○	○	○	○

help with projects and goals.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Community interest groups (e.g., chamber of commerce, neighborhood associations, nonprofits) turn to the library for help with projects and goals.	○	○	○	○	○
Our library collects data on library activities and services to inform decision making and planning.	○	○	○	○	○
Our library practices organizational transparency.	○	○	○	○	○

Please use this space to tell us more about your statements above.

Interests & Activities

Interests & Activities

Considering the interests of your local user community, indicate your library's likely or current level of activity, on a scale ranging from "Unlikely" to become active to "Highly active" now.

	Unlikely	Somewhat likely	Likely	Somewhat active	Active	Highly Active
Provide links to open data portals or other open data sources through your website	○	○	○	○	○	○
Provide reference assistance for community members on finding or using open data	○	○	○	○	○	○
Build collections of open data of value to your community	○	○	○	○	○	○
Offer professional development to library personnel on open data	○	○	○	○	○	○

	Unlikely	Somewhat likely	Likely	Somewhat active	Active	Highly Active
Offer programs for the community on the benefits and use of open data	○	○	○	○	○	○
Collaborate with local government agencies to make their data available to the public	○	○	○	○	○	○
Publish our library's data as open data	○	○	○	○	○	○

Please use this space to tell us more about your activities above.

Capacity & Resources

Capacity & Resources

In relation to the open data interests and activities indicated on the previous page, how important are the following resources, on a scale from “Not important” to “Essential”?

	Not important	Low importance	Moderate importance	High importance	Essential
Professional development on open data access, use, and instruction	○	○	○	○	○
Professional development on preparing and publishing datasets	○	○	○	○	○
Professional development on curation and infrastructure for data collections	○	○	○	○	○
New public service position(s) with open data expertise	○	○	○	○	○
New technical position(s) with open data expertise	○	○	○	○	○

	Not important	Low importance	Moderate importance	High importance	Essential
Software and tools for preparing and publishing open data	○	○	○	○	○
Access to a platform and support for publishing open data	○	○	○	○	○
Improved library technology for hosting data	○	○	○	○	○
New or improved collaboration with city, state, or other partners	○	○	○	○	○
Grant funding for development and implementation of open data activities	○	○	○	○	○
Increased operating budget	○	○	○	○	○

Please use this space to tell us more about your answers above.

Alignment

Alignment

Please indicate the degree to which you agree with the following statements.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Not sure
Open data initiatives align with our library's mission.	○	○	○	○	○	○
Open data initiatives align with our community's interests.	○	○	○	○	○	○
Open data						

initiatives align with our library's current technical proficiency.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Not sure
It will be very difficult for our library to prioritize open data initiatives.	○	○	○	○	○	○
Progress with open data initiatives would require significant new funding.	○	○	○	○	○	○
Providing leadership for local open data initiatives should be a priority for our library.	○	○	○	○	○	○

Moving Forward with Open Data

Moving Forward with Open Data

Which types of open data education and support would your library be likely to take advantage of over the next three years?

- [ ] Online webinars
- [ ] Online curriculum
- [ ] Local or regional in-person workshops
- [ ] Short-term open data consultant
- [ ] Paid graduate student interns
- [ ] Unpaid student practicum or capstone projects
- [ ] Community volunteers

Which positions would be most likely to work on open data activities in your library?

What are your library's greatest barriers to becoming involved in open data?

Please use the space below to share additional comments. For example, you might comment on how open data fits into your library's priorities, or talk more about your answers to the questions above.

Optional Self-Identification

Optional Self-Identification

Providing this information is optional. If you provide your contact information, the ODL team may follow up with you.

If you do not wish to self-identify, please skip this section, and click the blue arrow at the bottom of this page to complete the survey and submit your responses.

Library Name: (Optional)

Your Name: (Optional)

Your Position: (Optional)

Email Address: (Optional)

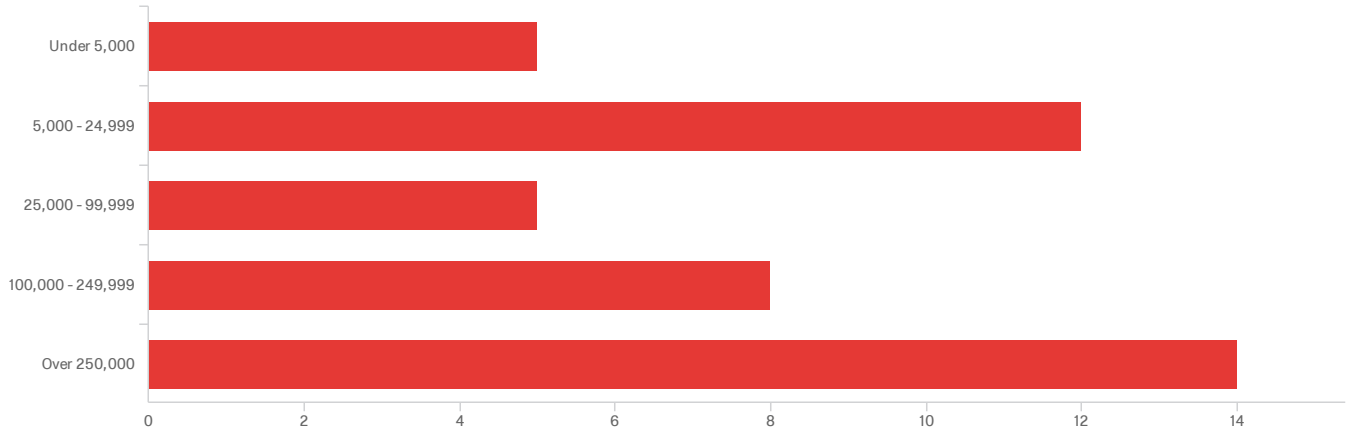
Powered by Qualtrics

Default Report

Public Libraries and Open Data Survey (Official Launch)

August 14, 2019 11:22 AM MDT

Service Population - Please select the size of your library's service population:

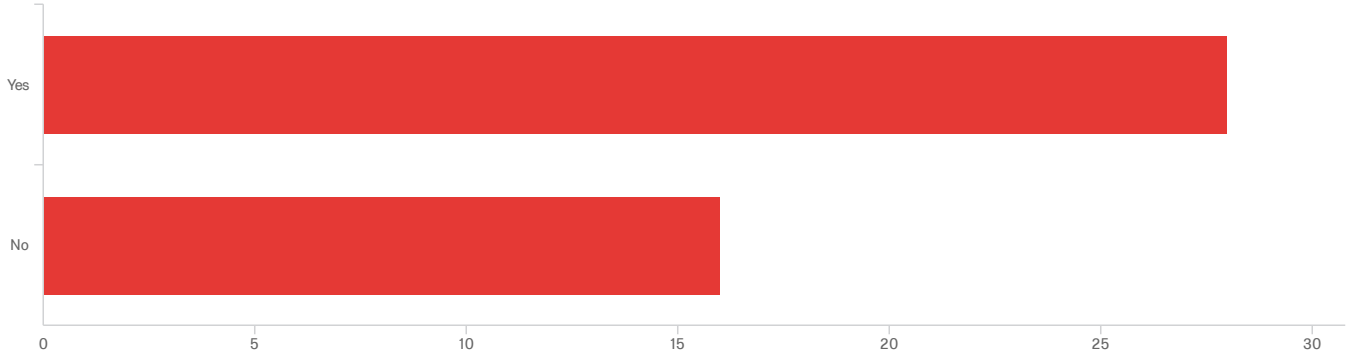


#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Please select the size of your library's service population:	1.00	5.00	3.32	1.44	2.08	44

#	Field		Choice Count
1	Under 5,000	11.36%	5
2	5,000 - 24,999	27.27%	12
3	25,000 - 99,999	11.36%	5
4	100,000 - 249,999	18.18%	8
5	Over 250,000	31.82%	14
			44

Showing rows 1 - 6 of 6

Library System Y/N - Are you representing a library system?



#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Are you representing a library system?	1.00	2.00	1.36	0.48	0.23	44

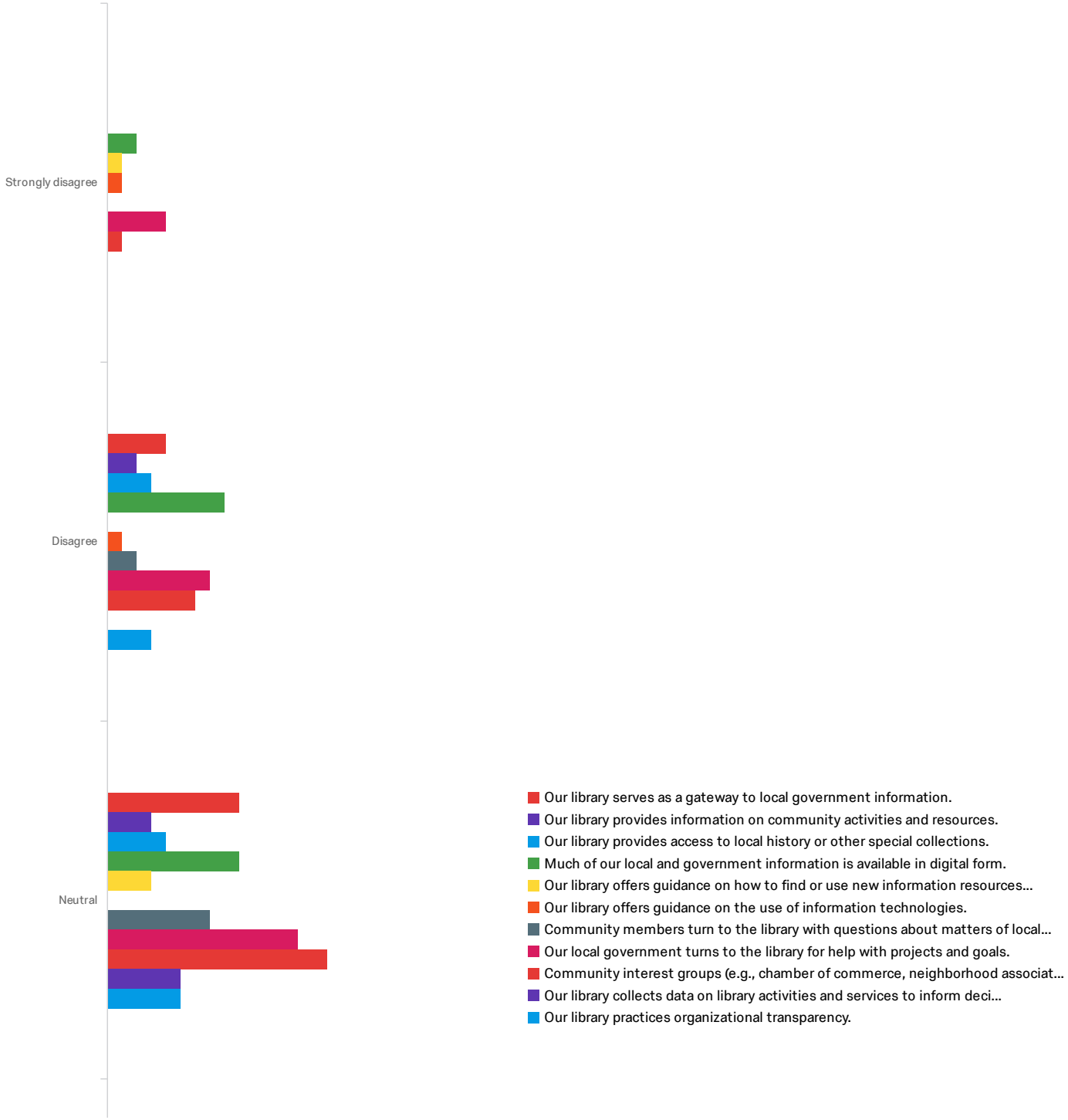
#	Field		Choice Count
1	Yes	63.64%	28
2	No	36.36%	16
			44

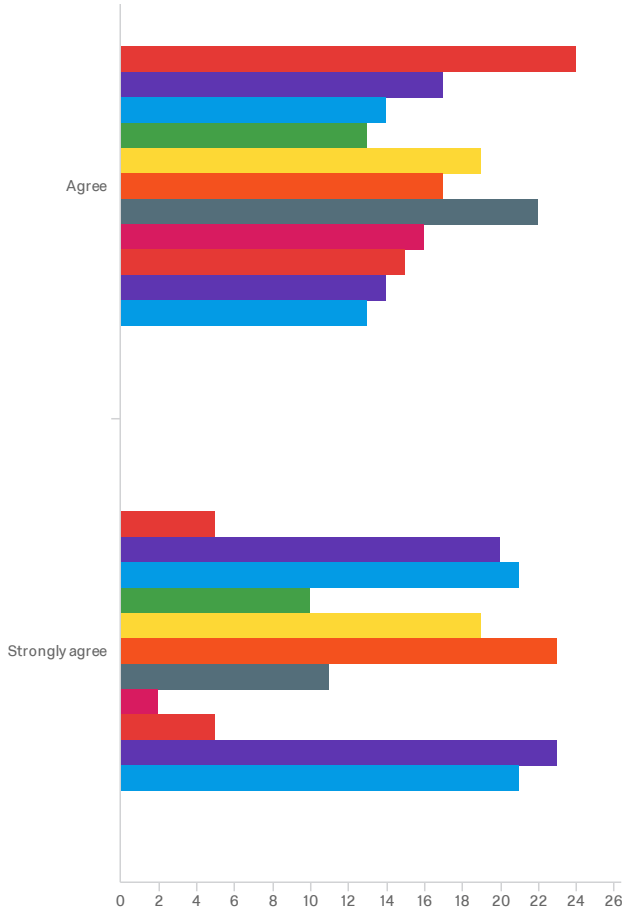
Showing rows 1 - 3 of 3

Extra org details - If you'd like to clarify any details about your service population or organizational structure, please do so below:

If you'd like to clarify any details about your service population or organ...
Part of a rural library district. With in town and rural Pierce County population with seasonal populations in the National Park and WSU students.
Branch manager of a larger county-wide system.
Library district currently only operates one library, so doesn't really qualify as a system of multiple libraries, but could at some point in the future. Also, did not want to imply that this was a city department library.
I am representing a branch of a library system
I represent a single branch in a library system.
rural county library district
We are an island library district. We have a significant number of seasonal residents.
We are small rural community located in western washington. We are a branch of Spokane County Library District.
Public library, part of a City
Countywide, rural system
500,000 over 4200 square miles
The Anacortes Public Library serves the only town/city on Fidalgo Island. There are some county residents that we serve as well.

Q1 - Current ContextOpen data activities relate to many established library roles, including reference service, provision of digital collections, and community programming and engagement. Please indicate the degree to which you agree with the following statements.





#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Our library serves as a gateway to local government information.	2.00	5.00	3.71	0.80	0.63	42
2	Our library provides information on community activities and resources.	2.00	5.00	4.31	0.80	0.64	42
3	Our library provides access to local history or other special collections.	2.00	5.00	4.26	0.90	0.81	42
4	Much of our local and government information is available in digital form.	1.00	5.00	3.50	1.18	1.39	42
5	Our library offers guidance on how to find or use new information resources.	1.00	5.00	4.31	0.80	0.64	42
6	Our library offers guidance on the use of information technologies.	1.00	5.00	4.43	0.82	0.67	42
7	Community members turn to the library with questions about matters of local public interest.	2.00	5.00	4.00	0.79	0.62	42
8	Our local government turns to the library for help with projects and goals.	1.00	5.00	3.12	1.05	1.10	42

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
9	Community interest groups (e.g., chamber of commerce, neighborhood associations, nonprofits) turn to the library for help with projects and goals.	1.00	5.00	3.40	0.95	0.91	42
10	Our library collects data on library activities and services to inform decision making and planning.	3.00	5.00	4.43	0.69	0.48	42
11	Our library practices organizational transparency.	2.00	5.00	4.24	0.92	0.85	42

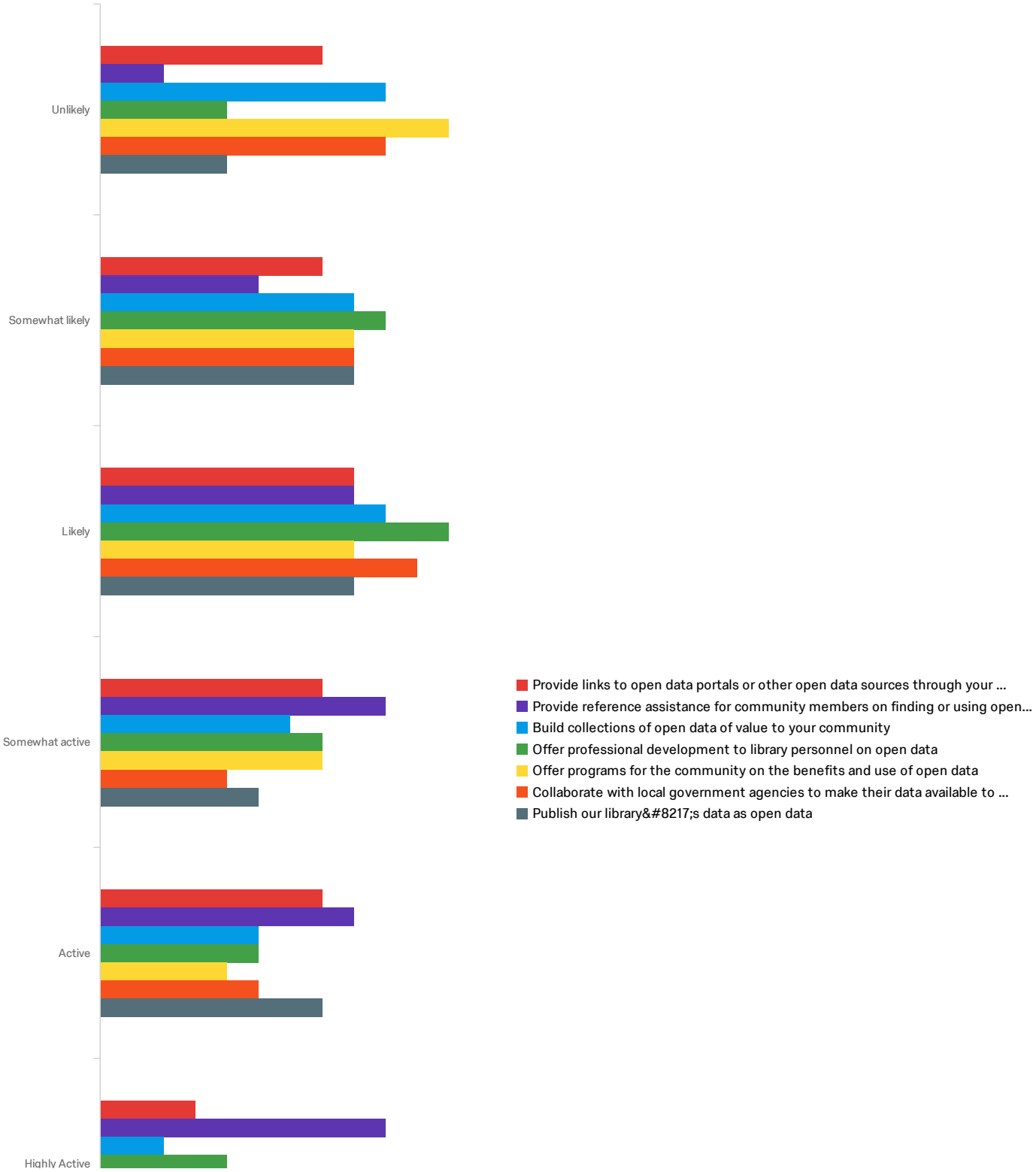
#	Field	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Total
1	Our library serves as a gateway to local government information.	0.00% 0	9.52% 4	21.43% 9	57.14% 24	11.90% 5	42
2	Our library provides information on community activities and resources.	0.00% 0	4.76% 2	7.14% 3	40.48% 17	47.62% 20	42
3	Our library provides access to local history or other special collections.	0.00% 0	7.14% 3	9.52% 4	33.33% 14	50.00% 21	42
4	Much of our local and government information is available in digital form.	4.76% 2	19.05% 8	21.43% 9	30.95% 13	23.81% 10	42
5	Our library offers guidance on how to find or use new information resources.	2.38% 1	0.00% 0	7.14% 3	45.24% 19	45.24% 19	42
6	Our library offers guidance on the use of information technologies.	2.38% 1	2.38% 1	0.00% 0	40.48% 17	54.76% 23	42
7	Community members turn to the library with questions about matters of local public interest.	0.00% 0	4.76% 2	16.67% 7	52.38% 22	26.19% 11	42
8	Our local government turns to the library for help with projects and goals.	9.52% 4	16.67% 7	30.95% 13	38.10% 16	4.76% 2	42
9	Community interest groups (e.g., chamber of commerce, neighborhood associations, nonprofits) turn to the library for help with projects and goals.	2.38% 1	14.29% 6	35.71% 15	35.71% 15	11.90% 5	42
10	Our library collects data on library activities and services to inform decision making and planning.	0.00% 0	0.00% 0	11.90% 5	33.33% 14	54.76% 23	42
11	Our library practices organizational transparency.	0.00% 0	7.14% 3	11.90% 5	30.95% 13	50.00% 21	42

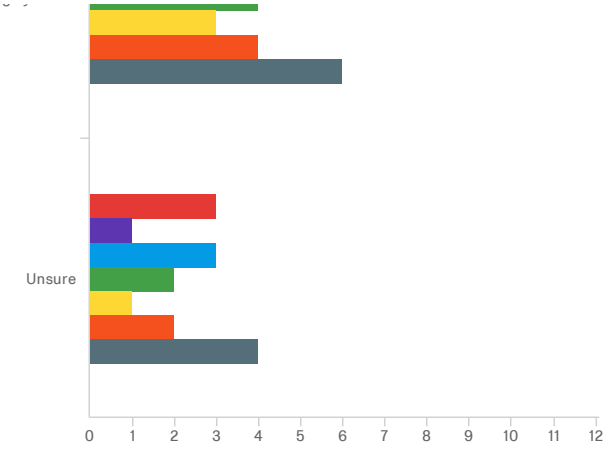
Showing rows 1 - 11 of 11

Q2 - Please use this space to tell us more about your statements above.

Please use this space to tell us more about your statements above.
While we are not consulted on goals and decisions from the town or Chamber of Commerce we are a source of information about the town goals. As Supervising Librarian I attend town and Chamber meetings so offer input in those open forums.
Our library system covers a county in where some city district keep their record digitized and some have not yet. Access to things of that nature will vary depending on the availability provided by the governing agency.
Our unique situation as a new district recently consolidating in a city library has created a situation conducive to working more closely with city government. However, county government tends not to reach out to this library district for help as the district is newly created and is still in growth phase.
Despite belonging to numerous local organizations, many of them still don't understand all the services we provide, both in print and online.
I feel like other organizations are unaware of what an amazing resource the public library can be for them and that they rarely think of us when they need help with something or to seek our support for their initiatives. This is an area for improvement that I hope to work towards in the coming years.
We could do more to communicate to local government and community interest groups that we are a resource for their projects and goals, and to actually do the work of being a resource.
Nothing to add
The local government does collaborate with us and rely on us to organize local service groups to help clean up the local park. The Chamber of Commerce rely upon us a LOT to pass on customer comments regarding what local residents want from their businesses. That success is due in part to my past service as a Board member of the Chamber. The comment that I hear the most is, the library appears to have their pulse on the community. Just this year, we have been invited to collaborate with the City, Fire Department, School District, and the Chamber of Commerce to plan and offer a family Christmas tree lighting ceremony and activities for families. I am currently involved and working with the Chamber of Commerce's event's committee and this year I have worked with the Chamber to offer a local Farmer's Market that has so far been well attended.
n/a
We are so small we seldom get questions like what are listed here.
There may be epistemic challenges to saying I "strongly agree" with the statement "Our library practices organizational transparency." We believe we are transparent, but how do we know? But sure, we are transparent to a degree.
Serving five counties, incorporated and unincorporated areas, presents challenges around the availability of local government information in certain areas.
Governmental and organizational cooperation varies by community.
The City of Anacortes has a City Museum which houses the core of local history, offers programs on history of the area, etc. The library's community room is the only meeting room. It is very popular with organizations because of the size, technology available, and no fee. It must be open to the public.

Q3 - Interests & Activities Considering the interests of your local user community, indicate your library's likely or current level of activity, on a scale ranging from "Unlikely" to become active to "Highly active" now.





#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Provide links to open data portals or other open data sources through your website	1.00	7.00	3.50	1.79	3.20	42
2	Provide reference assistance for community members on finding or using open data	1.00	7.00	4.12	1.53	2.34	42
3	Build collections of open data of value to your community	1.00	7.00	3.19	1.79	3.20	42
4	Offer professional development to library personnel on open data	1.00	7.00	3.48	1.62	2.63	42
5	Offer programs for the community on the benefits and use of open data	1.00	7.00	2.95	1.68	2.81	42
6	Collaborate with local government agencies to make their data available to the public	1.00	7.00	3.19	1.79	3.20	42
7	Publish our library's data as open data	1.00	7.00	3.88	1.84	3.39	42

#	Field	Unlikely	Somewhat likely	Likely	Somewhat active	Active	Highly Active	Unsure
1	Provide links to open data portals or other open data sources through your website	16.67% 7	16.67% 7	19.05% 8	16.67% 7	16.67% 7	7.14% 3	7.14% 3
2	Provide reference assistance for community members on finding or using open data	4.76% 2	11.90% 5	19.05% 8	21.43% 9	19.05% 8	21.43% 9	2.38% 1
3	Build collections of open data of value to your community	21.43% 9	19.05% 8	21.43% 9	14.29% 6	11.90% 5	4.76% 2	7.14% 3

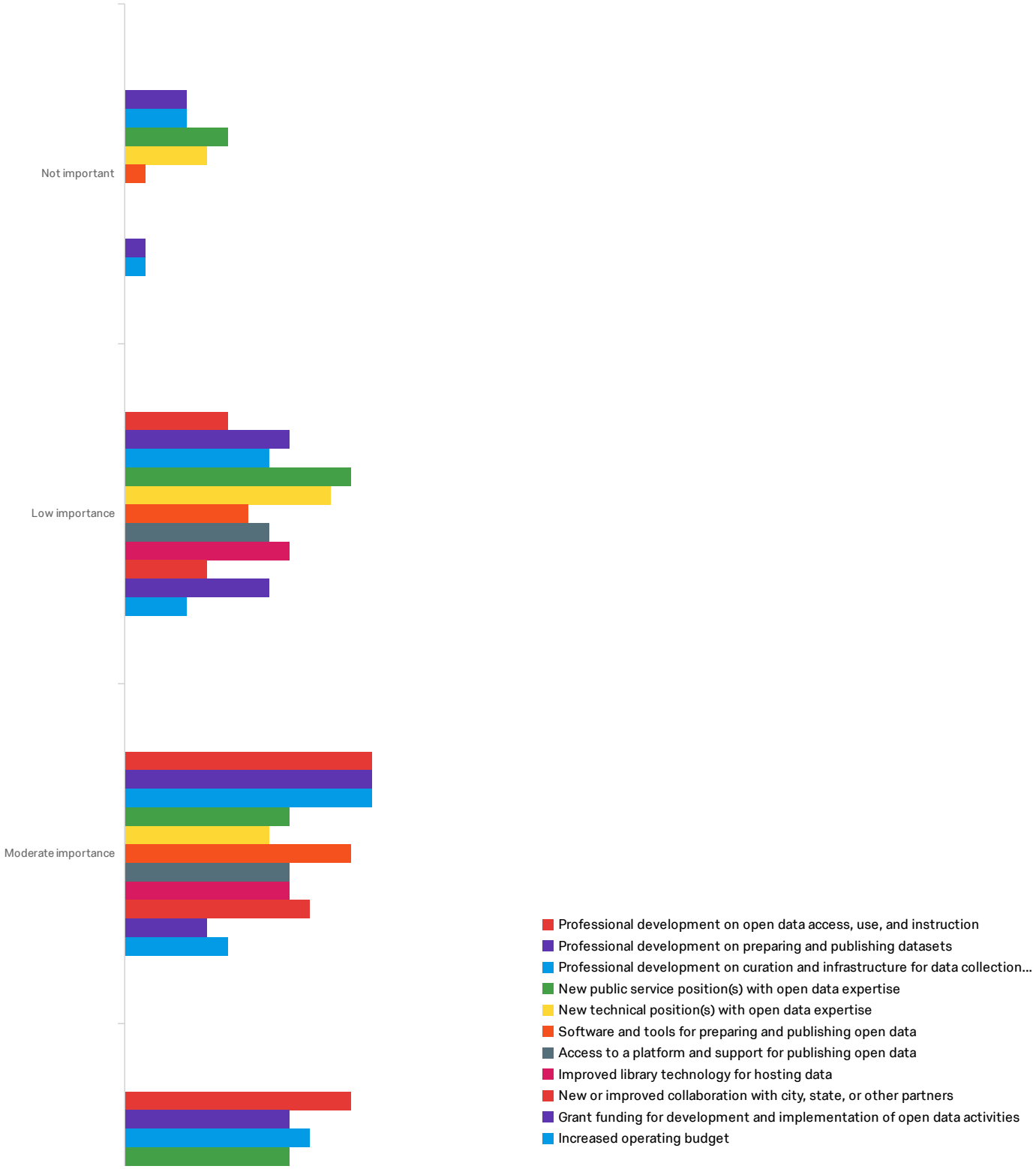
#	Field	Unlikely	Somewhat likely	Likely	Somewhat active	Active	Highly Active	Unsure
4	Offer professional development to library personnel on open data	9.52% 4	21.43% 9	26.19% 11	16.67% 7	11.90% 5	9.52% 4	4.76% 2
5	Offer programs for the community on the benefits and use of open data	26.19% 11	19.05% 8	19.05% 8	16.67% 7	9.52% 4	7.14% 3	2.38% 1
6	Collaborate with local government agencies to make their data available to the public	21.43% 9	19.05% 8	23.81% 10	9.52% 4	11.90% 5	9.52% 4	4.76% 2
7	Publish our library's data as open data	9.52% 4	19.05% 8	19.05% 8	11.90% 5	16.67% 7	14.29% 6	9.52% 4

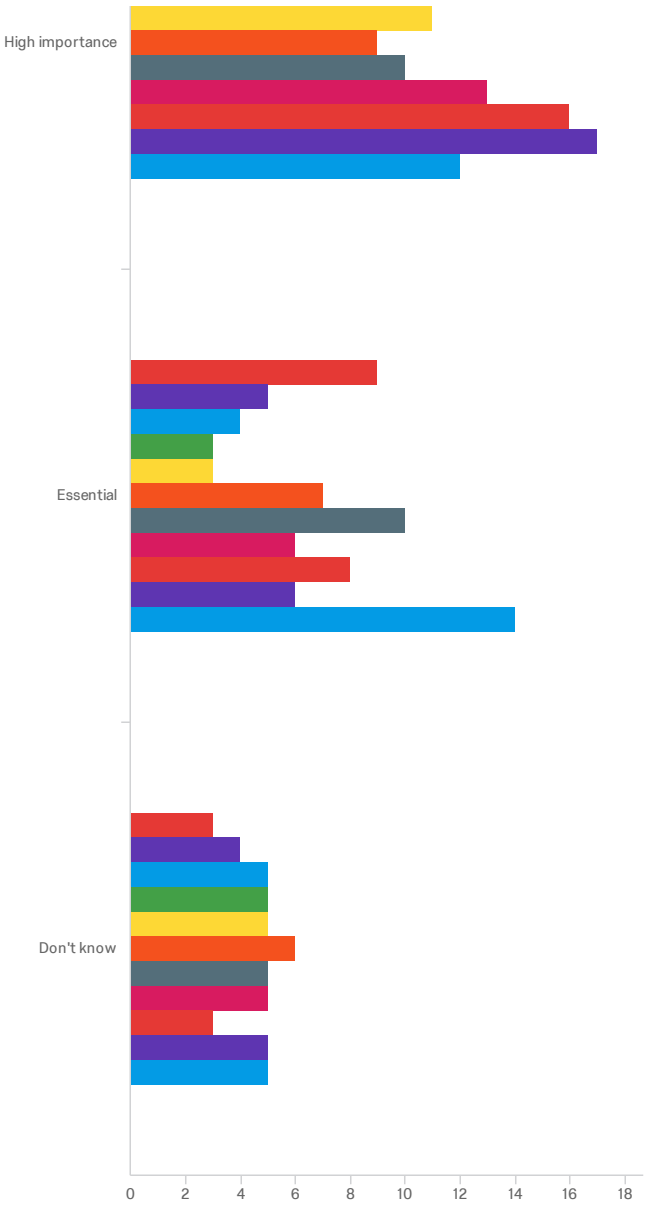
Showing rows 1 - 7 of 7

Q4 - Please use this space to tell us more about your activities above.

Please use this space to tell us more about your activities above.
The library offers access to a wide range of databases from the internet and purchased for us by our customers.
In an effort to 'de-clutter' the library's website, we removed many 'non-library' related links to local and social open data resources many in our community found helpful. Those links are instead bookmarks or otherwise saved on library computers and forward to interested patron when (if) requested. It takes some staff time to ensure the information is up-to-date and relevant to our patronage.
Small staff with wide range of technology skill will make this a challenging service without training and support.
nothing to add
Beyond any staff member's individual interest in the topic of open data, I don't believe that our library system, and definitely not my own branch, is doing much.
We have received several digitization grants (Rural Heritage Project) from the Washington State Library and those collections of local historical photos are available online.
We are currently hiring for a Community Data Coordinator at the library
I am very unsure what some of these question refer to. Sorry, we could clearly be involved in more of this than I am aware of. I am just unclear.
Our City already does a good job publishing GIS data. The library is currently constrained from offering new programs and services due to lack of staff.
Our priority is to first develop tools to use library data internally for decision-making. We would only move to publishing our library's data as open data after we meet our internal goals, due to limited resources and staffing for these tasks.
We are currently in the middle of a project with a ODL intern helping us to select and prepare local government datasets. We also participated in the Data Equity for Main Street project and have offered that curriculum to our public.
These are seldom activities we are asked to do.
We have published open data about our collection and collection use. We have also offered past workshops to the public on using open data.
Budget & Goal benchmarks are open data we participate in
Our Rural Library District has a very limited budget so we don't plan to pursue open data at this time / on our own. Should we partner or gain support from grants, other libraries or be approached by our local government entities, that could change.
Currently, we offer assistance with particular patron questions. We haven't developed plans to provide classes around open data or developed collections around open data. We have started this year to publish our library data through our open data portal.
The City has launched a broadband/fiber small business arm to the City's utilities programs and will be completed in 4 years. This fast fiber will be much more affordable than the current commercial offerings and this will help in providing equity in that inequity gap for those without internet service at home. 10% to 15% of the middle school and high school students do not have internet access at home and yet all their homework is online with use of their assigned school laptop.

Q5 - Capacity & Resources In relation to the open data interests and activities indicated on the previous page, how important are the following resources, on a scale from “Not important” to “Essential”?





#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Professional development on open data access, use, and instruction	2.00	6.00	3.83	1.14	1.29	40
2	Professional development on preparing and publishing datasets	1.00	6.00	3.40	1.39	1.94	40
3	Professional development on curation and infrastructure for data collections	1.00	6.00	3.48	1.41	2.00	40
4	New public service position(s) with open data expertise	1.00	6.00	3.20	1.54	2.36	40
5	New technical position(s) with open data expertise	1.00	6.00	3.35	1.49	2.23	40
6	Software and tools for preparing and publishing open data	1.00	6.00	3.83	1.36	1.84	40

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
7	Access to a platform and support for publishing open data	2.00	6.00	3.95	1.28	1.65	40
8	Improved library technology for hosting data	2.00	6.00	3.80	1.27	1.61	40
9	New or improved collaboration with city, state, or other partners	2.00	6.00	3.92	1.06	1.12	40
10	Grant funding for development and implementation of open data activities	1.00	6.00	3.88	1.29	1.66	40
11	Increased operating budget	1.00	6.00	4.25	1.20	1.44	40

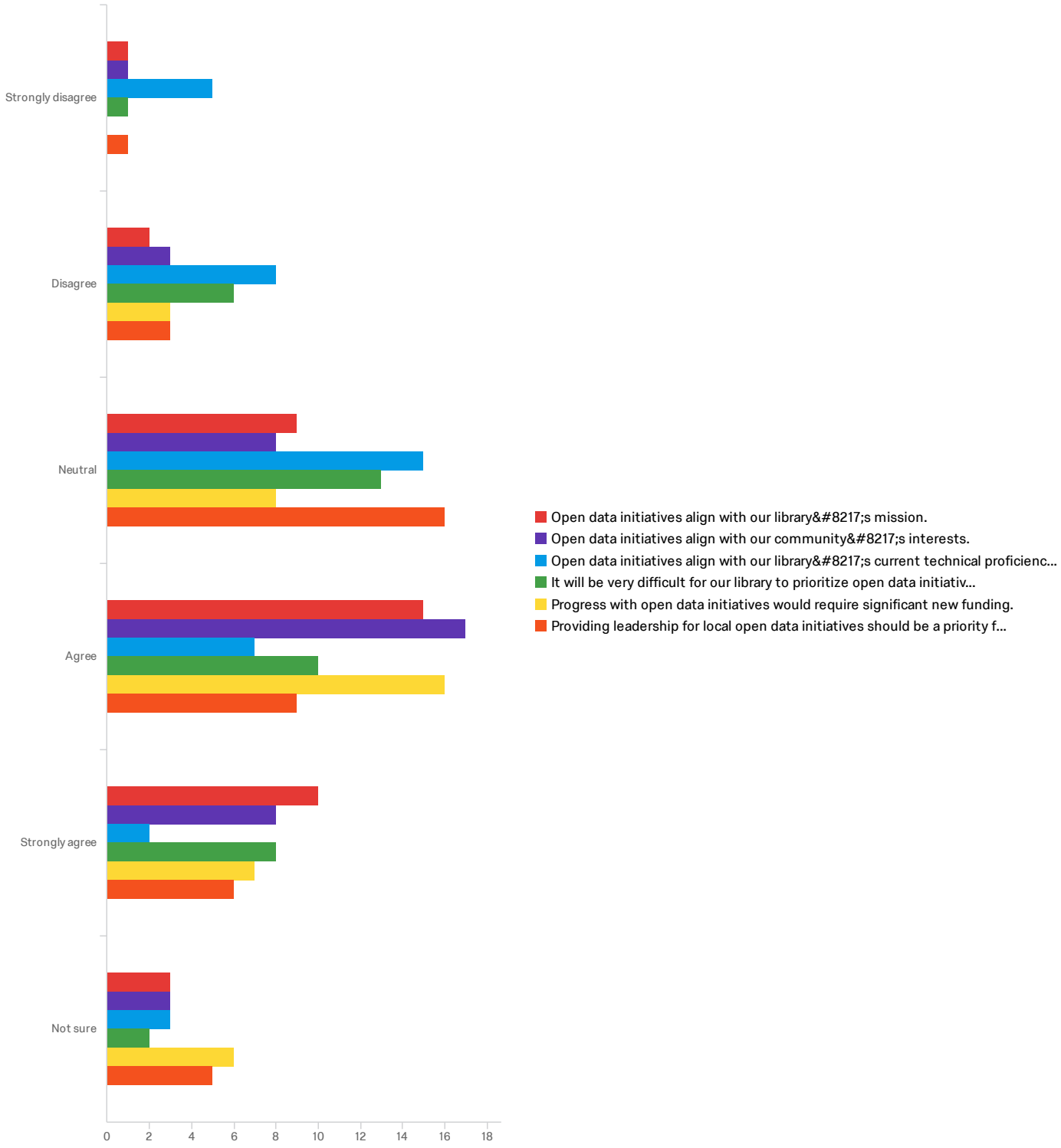
#	Field	Not important	Low importance	Moderate importance	High importance	Essential	Don't know	Total
1	Professional development on open data access, use, and instruction	0.00% 0	12.50% 5	30.00% 12	27.50% 11	22.50% 9	7.50% 3	40
2	Professional development on preparing and publishing datasets	7.50% 3	20.00% 8	30.00% 12	20.00% 8	12.50% 5	10.00% 4	40
3	Professional development on curation and infrastructure for data collections	7.50% 3	17.50% 7	30.00% 12	22.50% 9	10.00% 4	12.50% 5	40
4	New public service position(s) with open data expertise	12.50% 5	27.50% 11	20.00% 8	20.00% 8	7.50% 3	12.50% 5	40
5	New technical position(s) with open data expertise	10.00% 4	25.00% 10	17.50% 7	27.50% 11	7.50% 3	12.50% 5	40
6	Software and tools for preparing and publishing open data	2.50% 1	15.00% 6	27.50% 11	22.50% 9	17.50% 7	15.00% 6	40
7	Access to a platform and support for publishing open data	0.00% 0	17.50% 7	20.00% 8	25.00% 10	25.00% 10	12.50% 5	40
8	Improved library technology for hosting data	0.00% 0	20.00% 8	20.00% 8	32.50% 13	15.00% 6	12.50% 5	40
9	New or improved collaboration with city, state, or other partners	0.00% 0	10.00% 4	22.50% 9	40.00% 16	20.00% 8	7.50% 3	40
10	Grant funding for development and implementation of open data activities	2.50% 1	17.50% 7	10.00% 4	42.50% 17	15.00% 6	12.50% 5	40
11	Increased operating budget	2.50% 1	7.50% 3	12.50% 5	30.00% 12	35.00% 14	12.50% 5	40

Showing rows 1 - 11 of 11

Q6 - Please use this space to tell us more about your answers above.

Please use this space to tell us more about your answers above.
Developing funding for development is in the scope of the Administrative staff and not done at a branch level.
We need all these things as we don't have an active staff training, community collaboration, or department resourcing and/or outsourcing open data materials.
Current situation of this library in managing rapid change since consolidation moving forward to a new building will make any major services changes in what the library offers a challenge through 2020.
I think open data is a topic that is very soon going to be far more important to libraries, even if I am not sure exactly what that will look like. I think the level of involvement we need to have will depend on how the public incorporates use of open data in their own lives. I would respond much more strongly if I were involved with an academic library, where I see involvement with open data as much more critical right now.
?? Again, at my level, I was unsure of how to respond to these questions.
Open data isn't even on the list of the services this library would like to add if we had the funding for additional positions. We'd need to have years of a higher level of funding before open data would be seen as an appropriate additional service to provide.
We would need financial help to institute any of these activities.
We simply do not have the staff skill or funding capacity to deal with being an advocate for open data at this time.
A successful example highlighted in a library journal, with a recipe for replication--this is a great way to get us to consider expanding open data activities
Again, because our funds are so limited we've been focused upon other programming areas and providing more basic services. BUT - we never know which way trends will go, where funding and training become available to us and/or what our community might start to desire in services.
In recent years we have added a position with a focus on data, provided training, acquired new software tools, and started to integrate those tools with our data sources. Our capacity to provide increased support for open data depends on having the resources available.
We are just gearing up to respond in the fashion to these questions above. Very exciting to launch much of this. Need funding to do much of it. Looking forward to exploring alternate funding at the state and federal level.

Q7 - Alignment Please indicate the degree to which you agree with the following statements.



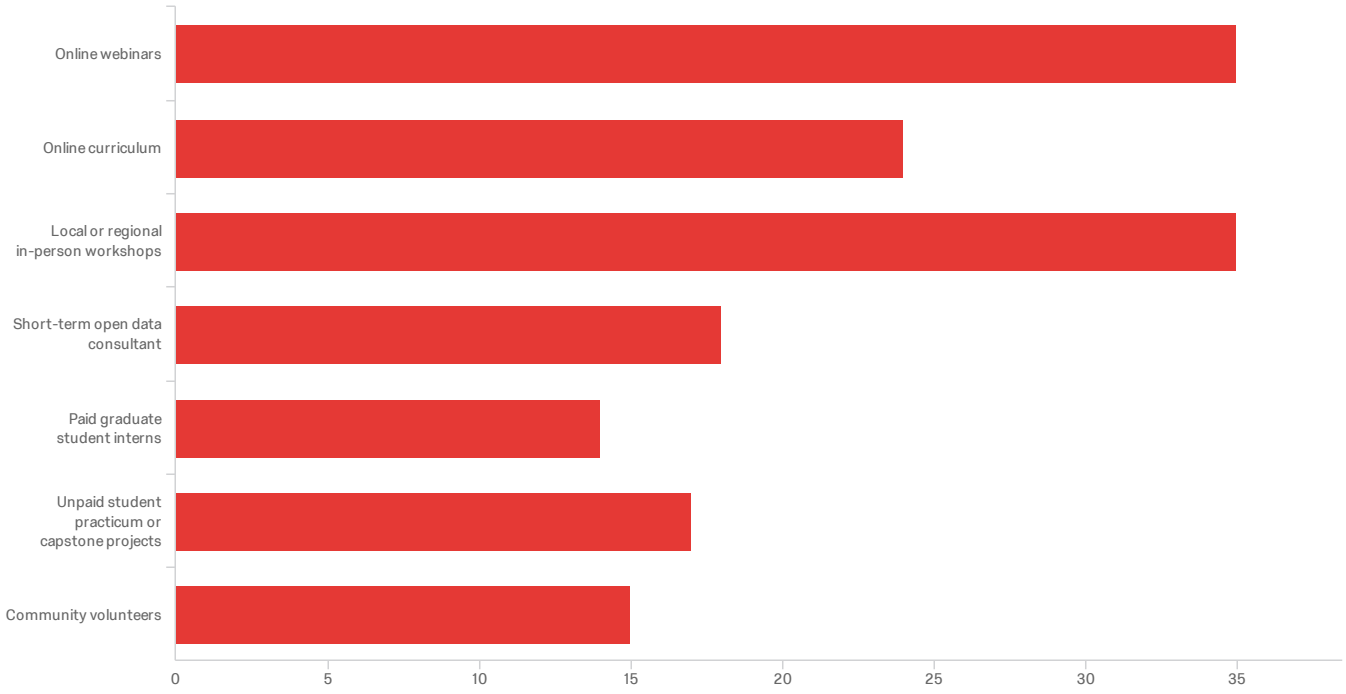
#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Open data initiatives align with our library's mission.	1.00	6.00	4.00	1.10	1.20	40
2	Open data initiatives align with our community's interests.	1.00	6.00	3.92	1.10	1.22	40
3	Open data initiatives align with our library's current technical proficiency.	1.00	6.00	3.05	1.32	1.75	40
4	It will be very difficult for our library to prioritize open data initiatives.	1.00	6.00	3.60	1.18	1.39	40
5	Progress with open data initiatives would require significant new funding.	2.00	6.00	4.13	1.12	1.26	40
6	Providing leadership for local open data initiatives should be a priority for our library.	1.00	6.00	3.77	1.23	1.52	40

#	Field	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Not sure	Total
1	Open data initiatives align with our library's mission.	2.50% 1	5.00% 2	22.50% 9	37.50% 15	25.00% 10	7.50% 3	40
2	Open data initiatives align with our community's interests.	2.50% 1	7.50% 3	20.00% 8	42.50% 17	20.00% 8	7.50% 3	40
3	Open data initiatives align with our library's current technical proficiency.	12.50% 5	20.00% 8	37.50% 15	17.50% 7	5.00% 2	7.50% 3	40
4	It will be very difficult for our library to prioritize open data initiatives.	2.50% 1	15.00% 6	32.50% 13	25.00% 10	20.00% 8	5.00% 2	40
5	Progress with open data initiatives would require significant new funding.	0.00% 0	7.50% 3	20.00% 8	40.00% 16	17.50% 7	15.00% 6	40
6	Providing leadership for local open data initiatives should be a priority for our library.	2.50% 1	7.50% 3	40.00% 16	22.50% 9	15.00% 6	12.50% 5	40

Showing rows 1 - 6 of 6

Q8 - Moving Forward with Open Data Which types of open data education and support would your library be likely to take advantage of over the next three years?



#	Field	Choice Count	
1	Online webinars	22.15%	35
2	Online curriculum	15.19%	24
3	Local or regional in-person workshops	22.15%	35
4	Short-term open data consultant	11.39%	18
5	Paid graduate student interns	8.86%	14
6	Unpaid student practicum or capstone projects	10.76%	17
7	Community volunteers	9.49%	15
			158

Showing rows 1 - 8 of 8

Q9 - Which positions would be most likely to work on open data activities in your library?

Which positions would be most likely to work on open data activities in you...
Education and Enjoyment Managers and IT manager and specialists
Librarians, Reference Associates
Paid interns?
public services staff and library administration, no IT staffing
library director
Public Services Assistants, Public Services Managers, Tech Support
IT Services Manager; ILS Coordinator; Online Services Coordinator
Public Service Librarians
IT and tech serv
Collection Services specialists
Library administration and digital services librarian
IT positions, Strategy Data Analysts
unknown at this time
Librarians
The Director and Assistant Director are the people on staff currently focused on open data with much of the work being handled by a paid intern.
Librarians
Director and IT staff
Director
Reference librarians
Data Librarian or Data Warehouse Administrator
Technology & Collections Manager & Outreach Coordinator
Librarians, Business Analyst, various IT, administrative assistants

Which positions would be most likely to work on open data activities in you...
Librarians and Public Service Assistants
Not sure - we'd have to restructure someone's responsibilities.
Our local history library assistant who works 5 hours a week on various research and digitization projects is currently looking into starting an open data partnership with local county agencies with assistance from WSL to aid her research.
Reference librarian
Library Assistants, Associates, and Librarians
Assistant Librarian
Compiling community information to support the work of partner agencies
This would be the major question. Probably our Adult Services Librarian in conjunction with our IT department that is city operated, not run by the library. So library expertise and IT do not collide here.
Adult Services staff, Director, or Technology Specialist
Technology Librarian Technology Assistant other Librarians and Library Assistants

Q10 - What are your library’s greatest barriers to becoming involved in open data?

What are your library’s greatest barriers to becoming involved in open data...
Funding
Technical know how, balancing confidentiality
time, staffing, philosophy, funding, everything as we have a lot going on already
Small staff with wide range of technical skill level. Current staff workloads are such that adding new initiatives in the near term would be almost impossible.
We're only opened 8 hours a week. Nobody has the time for this.
Lack of knowledge and application to our current work
Staff time and expertise, competing priorities, funding for additional staff
currently none
Not much current familiarity with this or other projects.
knowledge and prioritzation
Funding/staff level.
Staffing resources
Other activities align more directly with our organizational priorities at this time. In addition, significant work would be necessary to make our data usable, and we are not staffed for that effort.
organizational change
budget, community priorities, local gov't interest
We are already involved in open data.
We are focused on what community is asking for, they are not asking for open data
Need, skills and money
Staff skill, staff time, funding
Staff time and organizational support (lib is part of a larger organization)
1. Lack of priority among many competing projects. 2. Spotting unique opportunities in the community (not overlapping other organization's initiatives)

What are your library's greatest barriers to becoming involved in open data...
Time & money; further software for other types of open data publishing beyond what is provided by our city
As a district we have so many jurisdictions to serve. Most are not working in this area.
competeing priorities, and staff levels
Time, funding, priorities.
Staffing, budget, getting buy-in from other local agencies re: making their data accessible
Time and budget
Focus on other priorities may not leave us with resources to focus heavily on open data.
Community interest is low and getting permissions from partner agency can be cumbersome, esp. if expense is involved.
time, money, capacity
Time, capacity, money, training, everything.
Staff knowledge - why open data is important, how to set it up AND partnership with government to implement
Funding and Staff capacity to implement new resources, services, programs, staff and public training, etc.

Q11 - Please use the space below to share additional comments. For example, you might comment on how open data fits into your library's priorities, or talk more about your answers to the questions above.

Please use the space below to share additional comments. For example, you m...
I don't think it doesn't fit our philosophy per-say but it isn't a priority right now. Maybe once we get other things off our plate, we can redirect focus to these kinds of endeavors.
I'm not even sure I know what open data is.
You asked to send this to our Branch Managers; only 2 of 10 have MLIS degrees. I am not certain they all know what open data is well enough to answer your questions. They do great public service work but it's usually focused on the day-to-day and not bigger picture. As for our management team, we have a very ambitious strategic plan that does not currently have any open data-focused initiatives on it.
We're currently developing a new strategic plan, so I can't currently say how Open Data fits into our service plans. My guess is that it doesn't at all. We still struggle to cover desk hours and pull holds in less than a week, so any new funding will be used to address basic shortfalls before expanding to optional services. If it was something as simple as 'run this script to collect this data and send it to a statewide portal for publishing' then we'd probably be able to participate. But developing open data sharing solutions on our own isn't a priority for us.
Open data is not seen as a priority for our library
We are involved in municipal open data. We also use our internal data assets to make decisions. I can imagine greater participation in community open data. Not sure what this would look like exactly.
We're in the process of developing a new Strategic Plan for 2020-2022. Much of our focus for the next three years will be guided by that plan.
We would be interested in learning and participating with an open data project but are neophytes. This is a stand alone library with a maxed out staff.
This is crucial to our library right now and with the launch of municipal fiber project. Our librarians need to design and develop new services around the expanded internet access, such as digital resources, digital learning, build app resources in the community and partner with other library agencies and schools that expand our reach. Funding is a big issue. We have the will but need staffing and the equipment now. We are getting the staff in positions to respond, get them trained but they are still involved with other tasks that take their tame as well. We are streamlining and prioritizing as necessary but with 54% senior population the retirees have opinions on these developments.

Opt-in Library - Library Name: (Optional)

Library Name: (Optional)
Pierce county Library
Whatcom County Library System
Whatcom County Library System
Whatcom County Library System
Spokane Public Library
Orcas Island Library
Mid-Columbia Libraries
Everett Public Library
King County Library System
Asotin County Library
Camas Public Library
ling county librray system
Whitman County Rural Library District
Burlington Public Libraru
Timberland Regional Library
Grandview Library
Fort Vancouver Regional Library District
Puyallup
Anacortes Public Library

Opt-in position - Your Position: (Optional)

Your Position: (Optional)
Branch Manager
Executive Director
Library Director
Branch Manager
Assistant Director
Organizational Performance Project Manager
Director
Techology & Collections Manager
regional manager
Director
Director
District Manager - Circulation, Data and Assessment
Library Director
Executive Director
Director
Library Director

10 Recommendations for Creating Internships with the Public Sector

Bree Norlander

Immediately upon graduating with my MLIS, I began project-managing the IMLS grant funded Open Data Literacy project headed by Carole Palmer and Nicholas Weber. Now, three years into my role, we are wrapping up the project and reflecting on our progress. We have provided placements at six governmental or public library locations for fourteen students. We have solicited and co-created internship projects from our civic partners that can be completed in a limited timeframe with specific deliverables. As is often the case, the internship projects and experiences have improved over the years and I wanted to take this opportunity to offer guidance to anyone taking on the task of partnering with external organizations to provide internships that benefit all parties involved. To that end, I have distilled my thoughts into a top ten list of recommendations for creating successful internships in partnership with public sector organizations.

1. **Pay interns competitive wages**
   I have no doubt that offering our interns competitive wages during their approximately 8-week-long internships was one key to success. In our case, the stipends for internships were proposed in and paid for by the LB21 IMLS grant. The public sector partners committed their time and a physical location for the interns, but did not have to support the interns wages. Competitive pay meant that we could attract top-notch students to the opportunities and we could compensate them appropriately for the high-level work they performed. Moreover, unpaid internships can reinforce societal inequities (https://www.linkedin.com/pulse/how-unpaid-internships-undermine-diversity-inclusion-daniels-mba-jd) and we embrace the University of Washington's commitment "to advance the values of diversity, equity and inclusion" (https://www.washington.edu/diversity/).

2. **Scope the project**
   This is so important and yet difficult to describe in a "how-to" format. In our experience, most projects needed to be reduced in scope from the original idea. This is what our process looked like when it was most successful:

1. We solicited ideas for internship projects from partner organizations 7 months prior to the internship start dates.
2. We chose to fund internships that were best aligned with our grant goals and had the most clarity in deliverables.
3. We spoke with the project partners to further refine the internship plan.
4. In collaboration with partners, we hired interns whose interests and experiences were well aligned with the specific projects.
5. We asked interns to meet with the partner and fill out a Plan of Work document (including sections for Problem Statement, Goals, Deliverables, Sustainability, Milestones). Within the goals section we included a section for “Out of Scope” and “Time Allowing.” These sections provided a placeholder for the many ideas that surface but aren’t necessarily accomplishable within the internship timeframe.
6. We reviewed the Plans of Work, provided feedback, and checked in throughout the internship on the progress.

A successfully scoped project feels like it has well-defined “walls” (what is within scope and what is without) and deliverables. It evokes a clear sense of steps required for completion. From day one, the intern will understand where to begin and how to progress. To gauge the scope, ask the intern, “do you know what to do next and what you will be doing for the next week?” If they are hesitant, the scoping needs more work.

3. **Flexibility begets success**

As the project gets underway, new information is likely to change the direction of the project or the specific deliverables. As an extreme example, one of our internships pivoted in its entirety when the COVID-19 pandemic hit Washington state and the sponsor organization, Washington State Library, needed help documenting public library service changes due to the pandemic. Be flexible with changes to the plan, but always make sure that the scoping is reasonable and that there are clear deliverables. Part of the learning that comes with any time-bounded project is adjusting expectations for what is possible, and what is feasible to actually complete in an 8 week internship. We have found that taking these adjustments seriously requires a project mentor and support team to be regularly available for consultation, and to be active in the revision of a project’s scope. Doing so teaches valuable project management skills to students and helps them quickly realize the realities of working in a dynamic public sector setting.

4. **Define roles and be available**

Too many cooks don’t necessarily spoil the broth as long as each has a defined role. These roles should be explained and agreed upon by all parties prior to the start of the internship. Ideally, hosting an internship onboarding event will allow you to cover this in detail. Our interns received mentoring from us (the academic institution) and our partners (public sector organizations and libraries). The academic partners can elicit feedback from interns on how the internship is going, provide literature recommendations, new ideas, administrative information or referrals, and

encouragement. The partners should provide the domain knowledge, orientation to the physical space, networking and introductions to other staff, and feedback on deliverables. We often found that one or two meetings that included all mentors and the intern were valuable in making sure everyone was on the same page.

5. **Create community**
   If you have multiple interns, even if they are working for different public sector partners, provide opportunities for them to work together. We created a Slack channel for each year's cohort and hosted happy hours and lunches throughout the internship timeframe. These opportunities allowed the interns to network, ask questions of each other, share ideas, and commiserate over shared frustrations.

6. **Provide opportunities for dissemination**
   Provide interns opportunities to present or publish the work they are doing. We created a Medium publication and had interns publish two or more blogposts during their internship which we promoted via social media. We also held a showcase event at the end of summer in which interns gave short presentations on their work. We shared invitations to this event with sponsors and the iSchool community at large. Events like the showcase or a poster session also provide opportunities to network, create artifacts for resumes and portfolios, and to immerse them in a community of domain enthusiasts. We've had interns submit journal articles and present at conferences based on their internship accomplishments. These experiences give students tangible experiences to use in job applications and interviews.

7. **Create an online workspace**
   Create an online place for the intern to store all relevant artifacts from datasets to workplans to blogpost drafts. We used Github repositories, but other platforms such as Open Science Framework (OSF), Gitlab, or Bitbucket are also available. These workspaces should be made accessible to all stakeholders of the internship, if not fully open to the public.

8. **Encourage documentation**
   The online workspace allows interns and stakeholders to track the progress of the internship. This is useful for following along and for retrospective review. Each internship will have different documentation requirements, but examples include methodologies (for data collection, for analysis, for tool building, etc.), code, changes to the Plan of Work, drafts of blogposts, slides for presentations. We often review our interns Github repositories when writing papers, presentations, and annual reports.

9. **Provide opportunities for remote work.**
   If at all possible, offer some remote internships. Our MLIS program offers both residential and online degrees and we strive to make opportunities available to all students. We worked closely with a couple partners to craft remote internships that mean we could recruit and hire some outstanding students who were not local. For us,

this required regular online check-ins via Zoom or other conferencing applications, but it was time well spent.

10. **Feed interns**
    Seriously, food goes a long way to creating community and building relationships. Host lunches, happy hours, or coffee times.

To celebrate a successful project, we are hosting an online alumni event for interns, public sector and public library sponsors, and our team. The enthusiasm for the event suggests to me that we made a lasting impact on interns and partners alike. Not only did the interns leave the experience with resume-building deliverables and experiences, but the partners continue to build upon and utilize deliverables. We, the ODL team, were able to use feedback from the interns about their experiences to update coursework to better prepare students for this type of work, and to create resources for professionals in the field to use now. Each year we improved our internship program based on the experiences of the previous year. In the end we have a cohort of former interns that are still excited about engaging and networking with each other, partners who are eager to hear from former interns, and new graduate students who are benefitting from a data curation curriculum that is fresh and relevant. I hope that others can benefit from what we learned as much as we did!

The Complicated Problem of Closing Open Data

Andrew Mckenna-Foster, University of Washington Information School

Kathleen Sullivan, Washington State Library, Office of Secretary of State

Where It All Began

A decade old, like many government data portals

Emphasis so far on participation and publishing

Appraisal and selection left to individual agencies

Information School
UNIVERSITY of WASHINGTON
ODL
OPEN DATA LITERACY
SOS
Office of the Secretary of State
Washington State Library

Good News, Bad News

Good: The portal is established, more data is open to the public, and use is accelerating.

Bad: Aging junk -- practice datasets, empty datasets, obsolete and low/no-use data.

Discovery gets harder. Keeping everything means the noise-to-signal ratio of searches will be high.

(Whyte & Wilson, 2010)

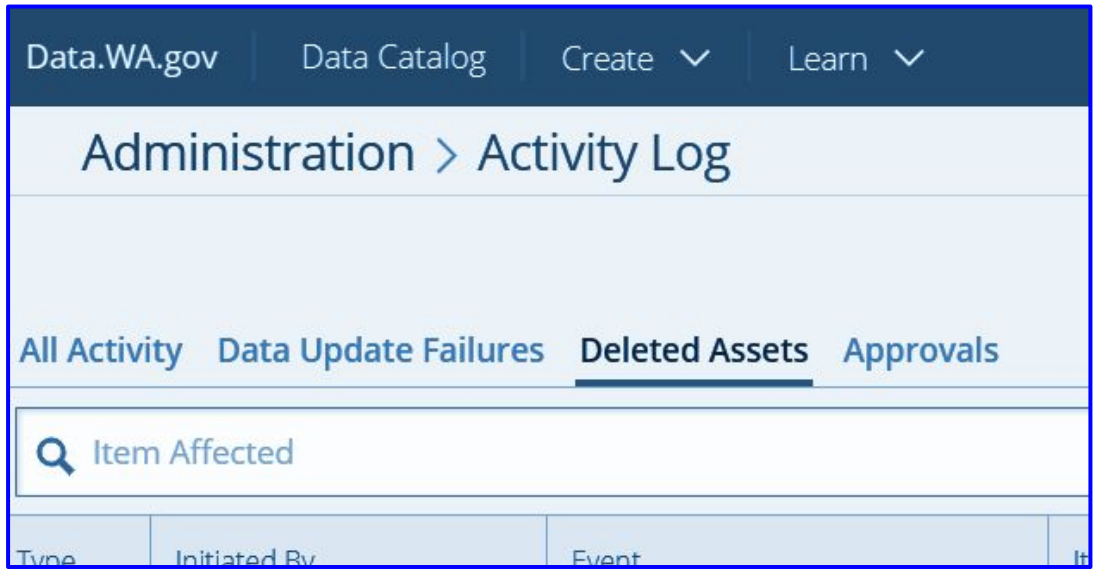
A partnership is born: OCIO + State Library

Libraries manage, prune collections

OCIO contracted with the Library to provide curation services -- prune “low-quality” data, better document and promote the rest, and support data quality going forward.

As a library intern, Andrew came in to conduct an initial assessment, but quickly recognized a snag …

How do you close open data (the right way)?



How do you un-select what was never (centrally) selected?

What is “low quality” data – inaccurate? obsolete? poorly documented? superseded?

How do you remove data in the spirit of transparency, access and public interest?

Time to Research: What We Did

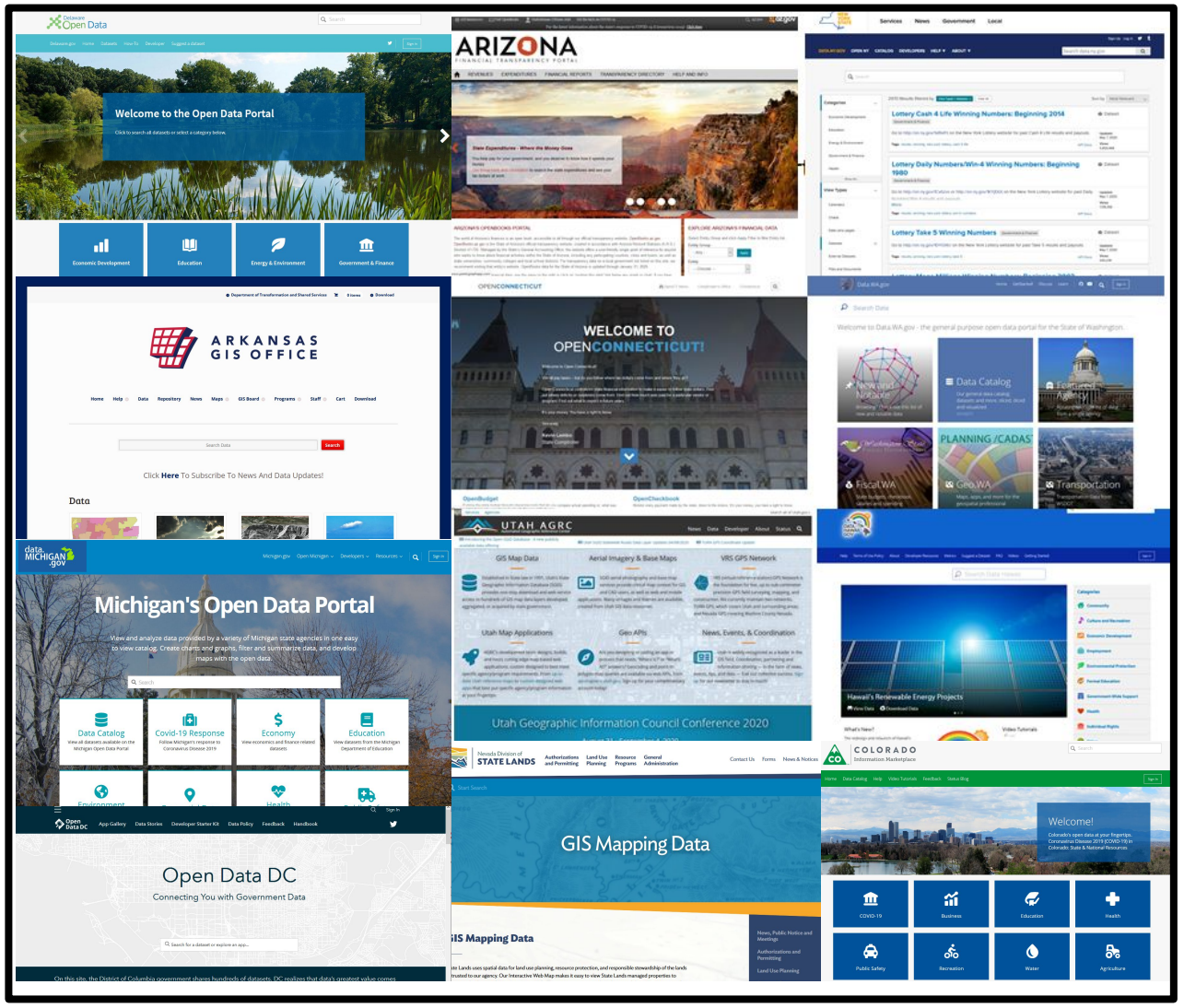
	B	C	D	E
	State	**URL**	**Date Visited**	**platform/source**
	Connecticut	https://www.osc.ct.g	20200113	Socrata
	Connecticut	https://data.ct.gov/	20200120	Socrata
	Delaware	https://data.delawar	20200122	Socrata
	Delaware	https://firstmap-dela	20200122	Esri
	District of Columbia	https://opendata.dc.	20200122	Esri
	Florida	http://geodata.myflo	20200124	Esri

Looked for data removal policies or practices at **other state government data portals**

Consulted **other domains** (e.g., archives, depositories) for weeding models

Portal Assessment

- 91 sites, 50 States:
  - 49 geospatial only
  - 13 transparency
  - 24 make tabular data available
  - 5 are other ‘open gov’ sites
- Sent 53 emails, 29 responses



Information School
UNIVERSITY of WASHINGTON

ODL
OPEN DATA LITERACY

SOS
Office of the Secretary of State
Washington State Library

data.ok.gov

search

About Datasets Snapshots

Home / Dataset

Type

Dataset

Categories

- Finance and Administration (100)
- Economic Development (46)
- Education (37)
- Health (35)
- Public Safety and Defense (19)
- Demographics (16)
- Natural Resources (11)
- Technology (11)
- Local Government (5)
- Social Services (4)
- Regulatory Services (3)
- Transportation (3)
- Other (2)
- Tourism And Recreation (1)

Topics

- economy (38)

Search

Search results

State-Managed Buildings Energy Savings Impact Dataset

Data reflecting the total energy saved in all state-managed buildings. Finance and Administration state buildings energy usage energy savings Field Value publisher-name data.ok.gov harvest_object_id 2ffz-i3xx harvest_source_id data.ok.gov harvest_last_upd

State Employee Turnover Rate Dataset

Maintain the state employee turnover rate at or below the annual regional average of surrounding states every year through 2019. Finance and Administration government state employee state personnel turnover Field Value publisher-name State of Oklahoma har

School Counselor- Student Ratio Dataset

Increase the school counselor to student ratio from 1 counselor per 421 students in 2013 to 1 counselor per 250 students by 2018. Education counselor school student ratio Field Value publisher-name data.ok.gov harvest_object_id 2jsz-jc8z harvest_source_id

Health Care Cost Growth Dataset

Limit state-purchased health care cost growth to 2% less than the projected national health expenditures average every year through 2019. Health healthcare health care cost growth Field Value publisher-name data.ok.gov harvest_object_id 2rx6-rfwj harvest_

AP Support Program (AVID) Dataset

Increase the percentage of public school districts participating in the Advanced Placement (AP) support program (AVID) from 3% in 2014 to 7% by 2018. Education ap advanced placement support program avid Field Value publisher-name data.ok.gov harvest_objec

Search

Categories

- Agriculture
- Consumer Protection
- Culture and Community
- Demographics
- Economics

Show All...

View Types

- Calendars
- Charts
- Data Lens pages
- Datasets
- External Datasets
- Files and Documents
- Filtered Views
- Forms
- Maps

Washington State Open Data

2879 Results

Sort by Most Relevant

Health Care Provider Credential Data

Health

Dataset

The Washington State Department of Health presents this information as a service to the public. True and correct copies of legal disciplinary actions taken after July 1998 are available on our Provider Credential Search site. These

More

Tags credential, provider, care, public, health, and 3 more

API Docs

Updated
May 10, 2020

Views
1,263,909

Elliot Bay

Natural Resources & Environment

Map

This is a map of cleanup sites and associated websites.

More

Tags ecology, cleanup status, cleanup, toxics cleanup, warm ranking, and 2 more

Updated
July 23, 2019

Views
216,436

Kenmore waterfront

Natural Resources & Environment

Map

This is a map of cleanup sites and associated websites.

More

Tags ecology, cleanup status, cleanup, toxics cleanup, warm ranking, and 2 more

Updated
July 23, 2019

Views
215,955

Find a Health Provider Credential

Health

Data Lens

This "data lens" page allows you to search the Department of Health's provider credential database for a specific

Updated

Few Policies Exist - Six States

5 List criteria

5 Publish procedures

4 Label retired datasets

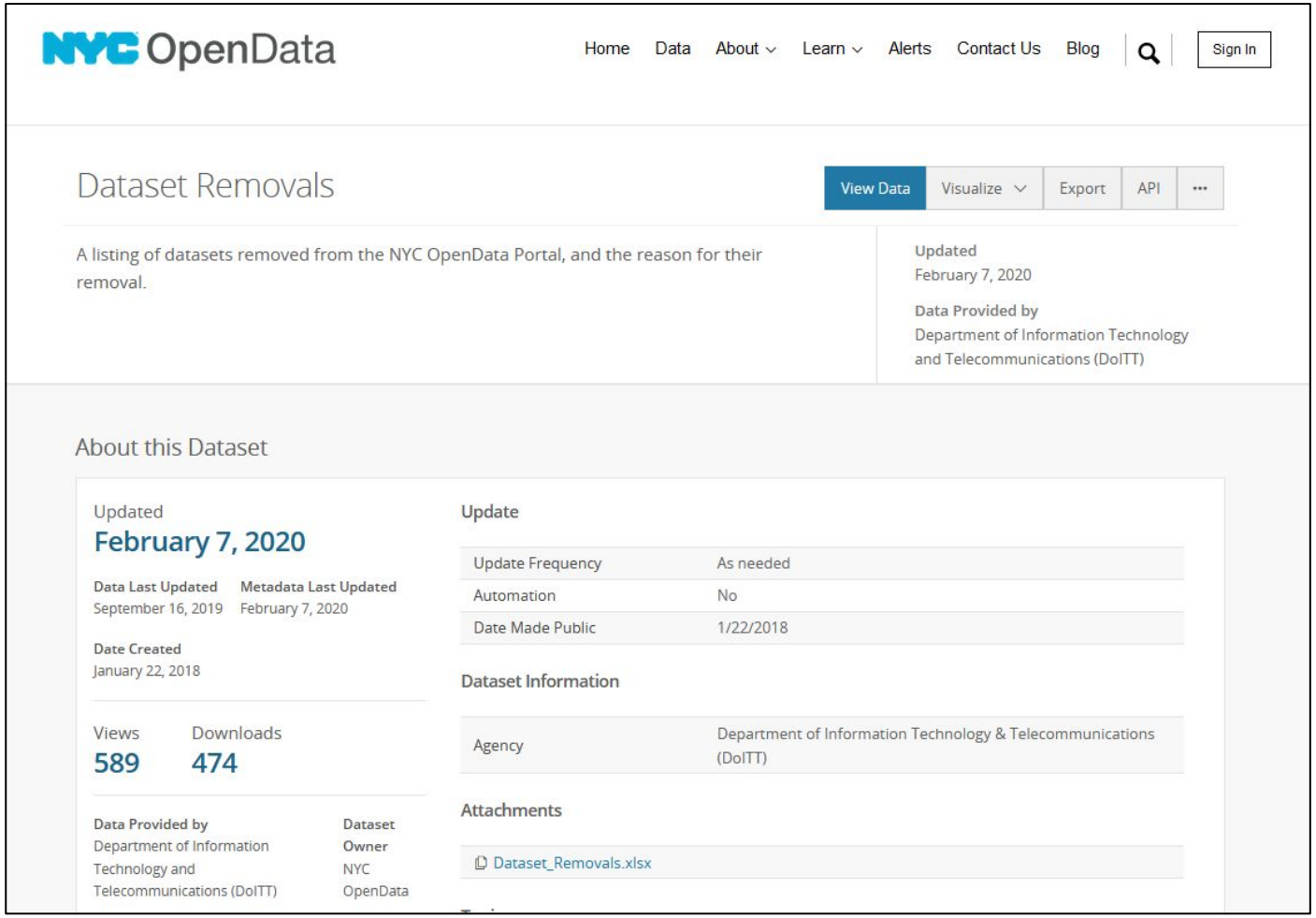
3 Use waiting periods

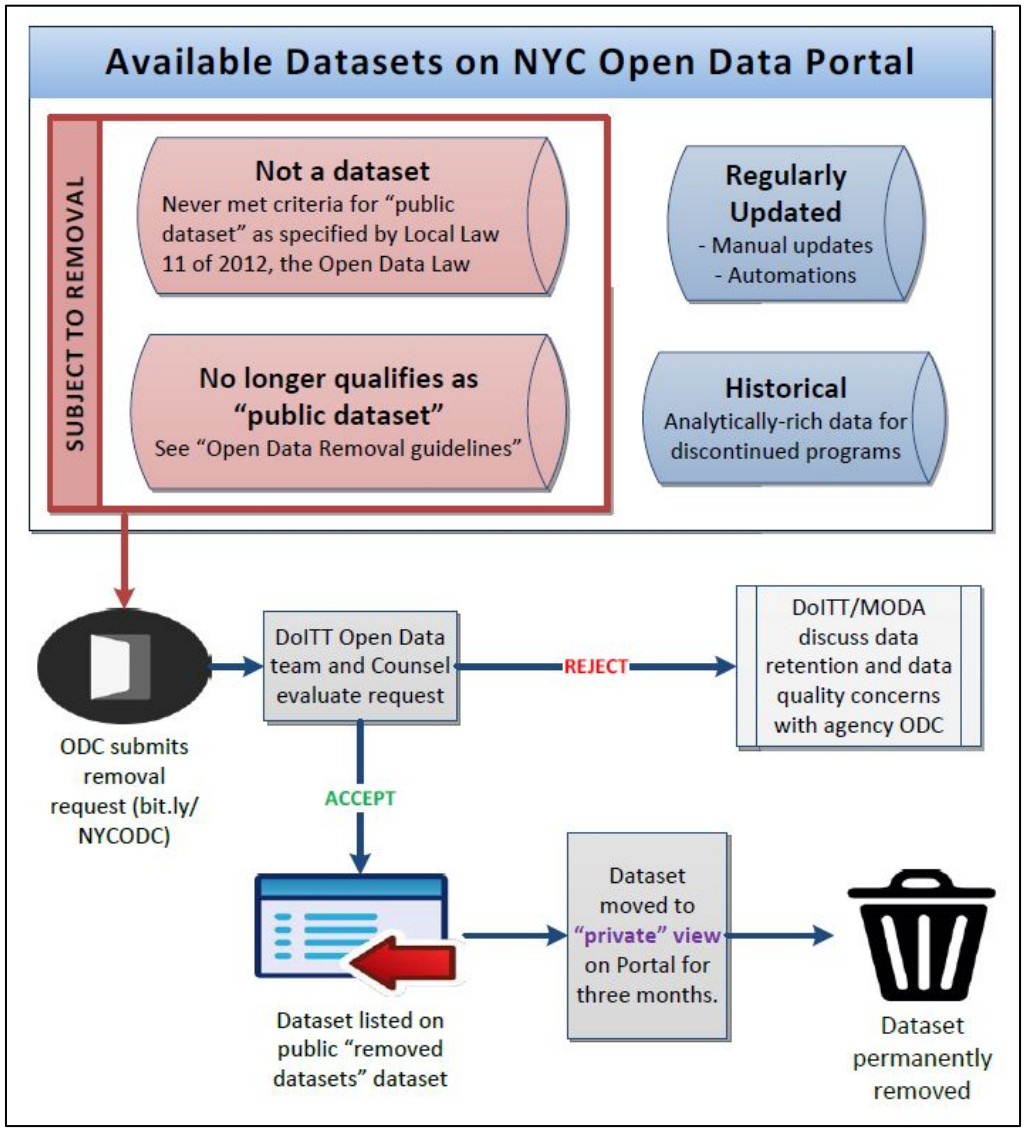
Information School
UNIVERSITY of WASHINGTON

ODL
OPEN DATA LITERACY

SOS
Office of the Secretary of State
Washington State Library

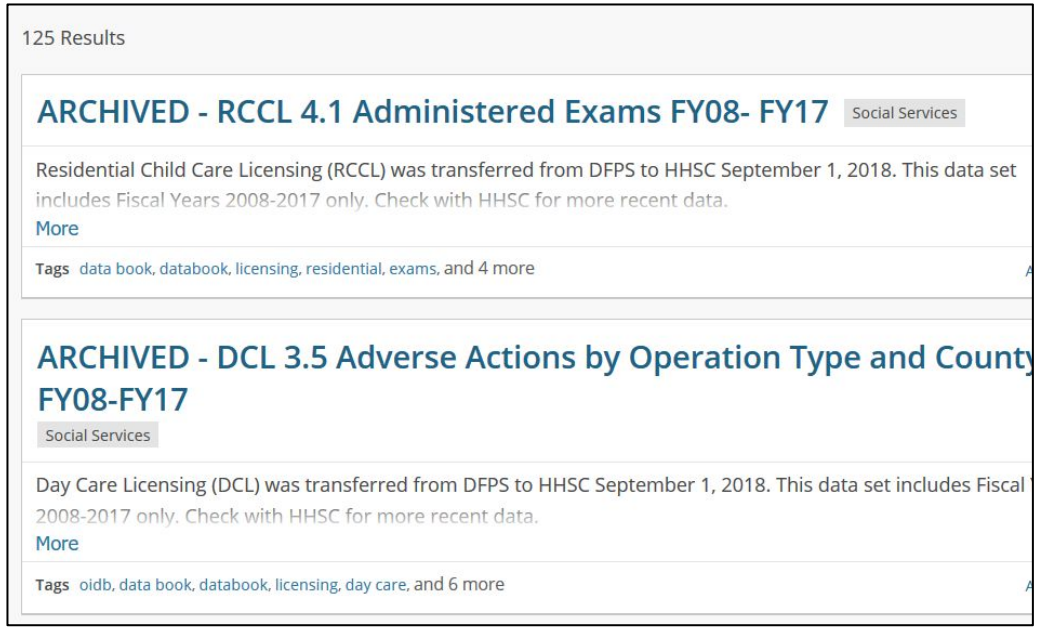
There is One Great Example

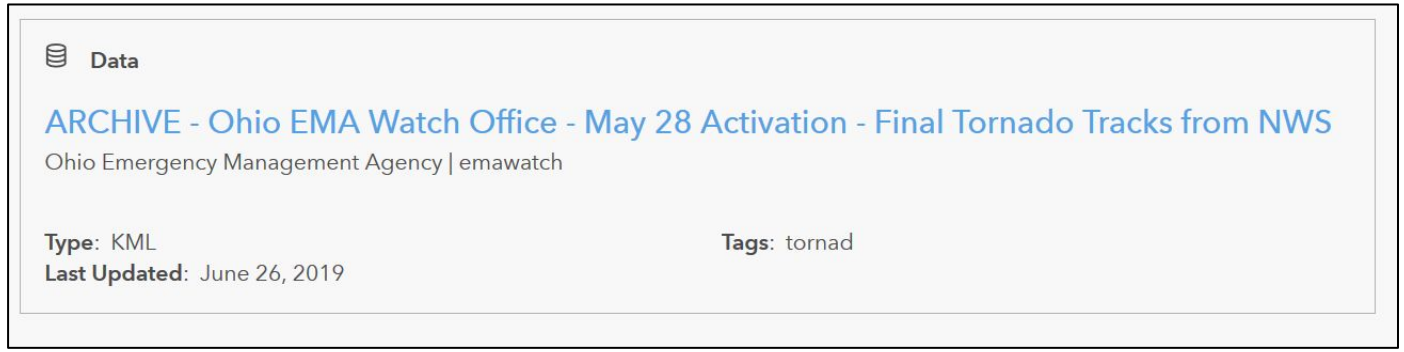


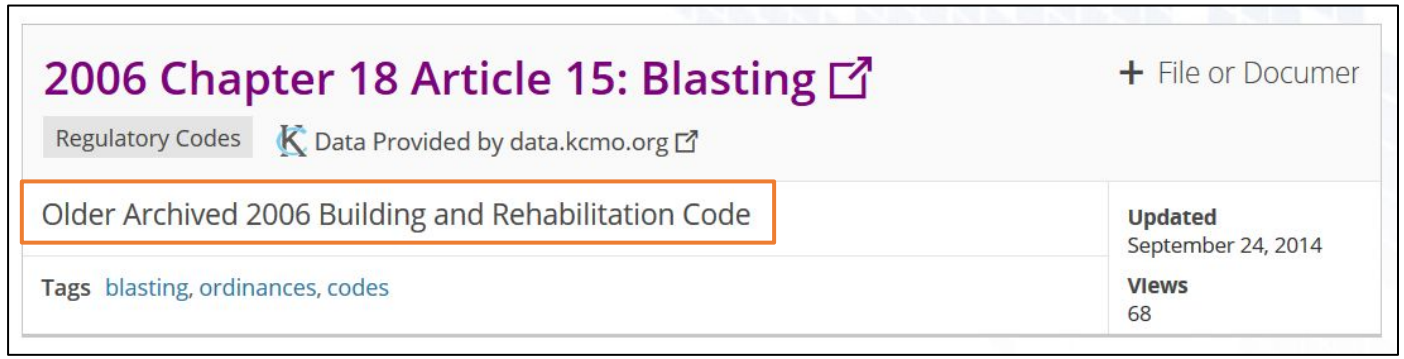


https://opendata.cityofnewyork.us/wp-content/uploads/2018/01/Open-Data-Removals-Process-and-Guidelines.pdf

Removal and Retirement is Occurring







Information School
UNIVERSITY of WASHINGTON
ODL
OPEN DATA LITERACY
SOS
Office of the Secretary of State
Washington State Library

Some things we noticed

- Difficult to understand who runs a portal
- Some portals are faceless

Photo by Fletcher6 - Own Work CC BY-SA 3.0

Data Provided by	Dataset Owner
(none)	Open Data Portal Kung-Fu Master

Subject: Re: A response was submitted through the form 'Contact Form'

Hello Kathleen,

Thank you for using [*state data portal*]. At this point we don't have policy/procedure for removing datasets. If any dataset needs to be removed, it will be decided by individual agency.

If we get more info, will keep you posted.

Thanks
[*state data portal*] Administrator

Portals with People, Policies

Retirement/Removal Models

Archival Record?

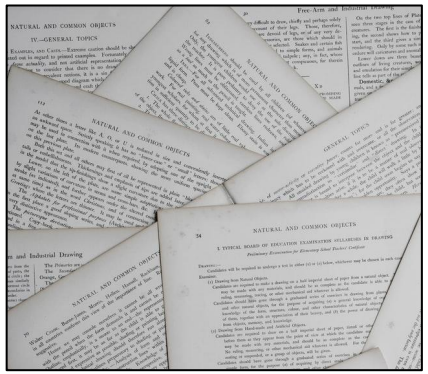


Photo by Annie Spratt on Unsplash

Government Document with Retention Schedule?

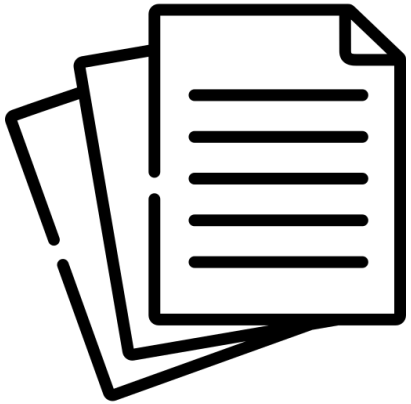
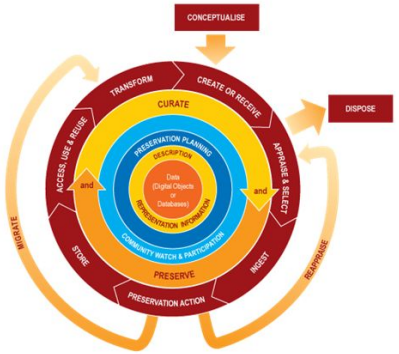


Image by Freepik on Flaticon

Resource in a Research Data Library?



Data Curation Centre

Resource in Traditional Library?

Photo by Rabie Madaci on Unsplash

Information School UNIVERSITY of WASHINGTON
ODL OPEN DATA LITERACY
SOS Office of the Secretary of State Washington State Library

Retirement/Removal Models

Archival Record?

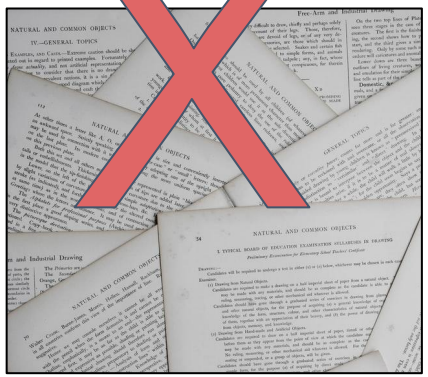


Photo by Annie Spratt on Unsplash

Government Document with Retention Schedule?

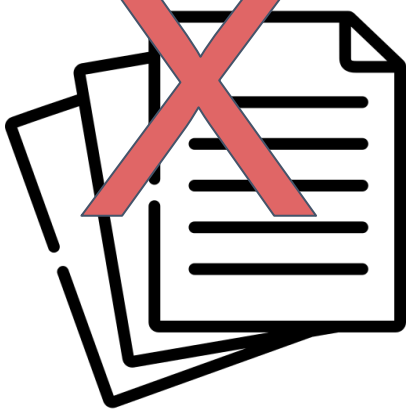
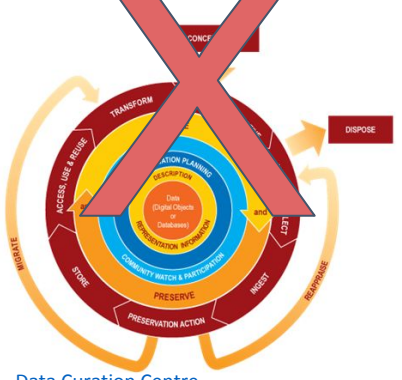


Image by Freepik on Flaticon

Resource in a Research Data Library?



Data Curation Centre

Resource in Traditional Library?

Photo by Rabie Madaci on Unsplash

Information School
UNIVERSITY of WASHINGTON

ODL
OPEN DATA LITERACY

SOS
Office of the Secretary of State
Washington State Library

Recommendations for data.wa.gov

Work with publishers and users on...

Appraisal and Selection Policy

Broad goals that work in a decentralized environment

Retirement Policy

Clear criteria
Procedures
Labeling

Connect with Digital Archives

about preserving historically valuable data

Make Portal Personable

Make policies and people easy to find

Questions?

Kathleen Sullivan, kathleen.sullivan@sos.wa.gov

Andrew Mckenna-Foster, andrew.mckennafoster@gmail.com

Information School
UNIVERSITY of WASHINGTON

ODL
OPEN DATA LITERACY

Office of the Secretary of State
Washington State Library

What data does your community need the most?

Washington State Library First Tuesdays Webinar
Tuesday, October 1
Presented by the Open Data Literacy team

ODL
OPEN DATA LITERACY

Information School
UNIVERSITY of WASHINGTON

Open Data Literacy:

Funded by IMLS, 2016-2020
Led by Carole Palmer and Nic Weber

INSTITUTE of Museum and Library SERVICES

Presenters:

- Carole Palmer
- Bree Norlander
- Kaitlin Throgmorton

ODL aims

Advance open data through education, research, and support for public librarians & information professionals in the public sector

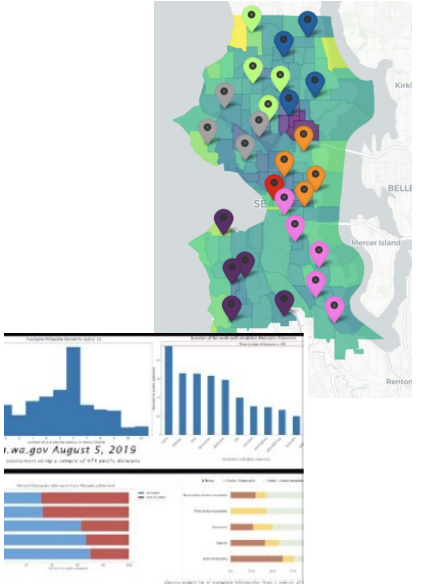
- new curriculum and educational resources
- field work experiences that benefit students, libraries, government
- collaboration with open data providers

Washington State · Office of the Chief Information Officer

Office of the Secretary of State
Washington State Library

Partners and students: 2019 internships

- Published library datasets
- Assessed dataset quality
- Transformed data for library applications
- Improved metadata
- Data management planning



Office of the Secretary of State
Washington State Library

The Seattle Public Library

WSDOT

City of Seattle

Asotin County Library

It's common to hear librarians lamenting the fact that they never get asked actual reference questions anymore. ...

But open data is an area where skilled guides are still needed. With the appropriate training, our librarians could be showing our community a richer way to view the world than through a simple google search.

I think that given the opportunity, many librarians would jump at the chance to learn more.

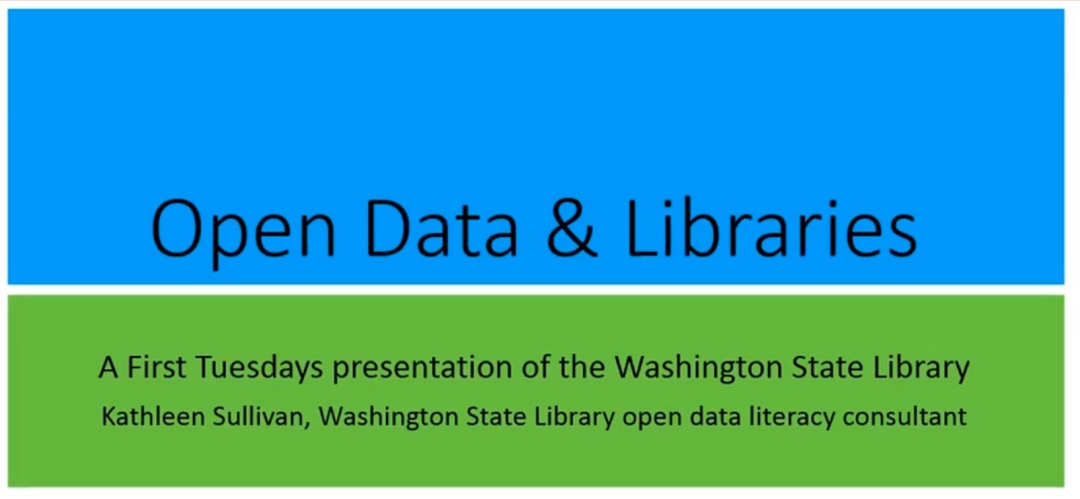
(Rural library district, service population 5,000 - 24,999)

Overview

- Building Upon “Why Open Data Belongs at Libraries”
- Tools to answer “What data does my community need?”
- Open Data Literacy Interns Address Library Data Needs
  - Asotin County Library
  - Seattle Public Library
- WA Libraries Involved in Open Data
  - Timberland Regional
  - Everett
- Statewide Survey
- Resources

Why Open Data Belongs at Libraries

First Tuesdays: December 4, 2018



https://my.nicheacademy.com/washingtonstate/course/8739

Brief Recap

- Open Data Activities: libraries publishing, teaching/reference, advising government
- What is Open Data?
  - “Open data and content can be freely used, modified, and shared by anyone for any purpose” - Open Knowledge International
- Data Portals
- Open data is used:
  - To tell a story
  - For efficiency and effectiveness of services
  - For innovation and entrepreneurship
- “Libraries help people access and use information”...“Open data fits snuggly into the libraries mission to empower people with information that helps improve their lives”

Administrative Boundaries | **Critical Areas Ordinance** | Environmental

- DOE_Aquifers
- Fish Distribution Data Description (PDF)
- Fish_Distribution
- Geographical Hazards Data Description (PDF)
- Geological Hazards
- HCA1_Major_Streams
- Kelp - Ealgrass Data Description (PDF)
- Kelp_Eelgrass
- Landslide Hazards Data Description (PDF)
- Landslide_Hazards
- Lummi Island Recharge 1978
- Major Streams Data Description (PDF)
- Marine Slope Stability Data Description (PDF)
- NRCA_Conservation
- Slope Stability Marine
- Species Habitat
- Species Habitats Data Description (PDF)
- Volcanic Hazards
- Volcanic Hazards Data Description (PDF)
- Wellhead Protection
- Wetlands
- Wetlands Data Description (PDF)
- Wildlife Corridor
- Wildlife Corridors Data Description (PDF)

Administrative Boundaries | Critical Areas Ordinance | **Environmental**

- 20 Foot Contours Data Description (PDF)
- 20 Foot Contours Data File (ZIP)
- Agriculture Protection Overlay Soils Data Description (PDF)
- Agriculture Protection Overlay Soils Data File (ZIP)
- Natural Resource Conservation Area Boundaries Data Description (PDF)
- Natural Resource Conservation Area Boundaries Data File (ZIP)
- Washington Water Resource Inventory Area Watershed Boundaries Data Description (PDF)
- Washington Water Resource Inventory Area Watershed Boundaries Data File (ZIP)

Administrative Boundaries | Critical Areas Ordinance | Environmental

- Cemetary_Districts
- County_Boundary
- Incorporated City
- Parks_Districts
- School Districts
- Zip Codes
- Zoning/Comprehensive Plan
- Fire Districts

https://www.whatcomcounty.us/716/Data

Analysis of Public Records Requests

https://sunlightfoundation.com/2018/10/16/results-from-analyzing-public-record-requests/

- “cities need to publish datasets that are of high interest to citizens – to get the benefits of open data and see a decrease in PRRs...We analyzed which types of information are the most popular to help cities do that.”
- 110,063 Public Records Requests from 33 cities

Analysis of Public Records Requests

https://sunlightfoundation.com/2018/10/16/results-from-analyzing-public-record-requests/

Most Popular Data Requests	
1	Police incident report
2	Parcel records, permits, plans
3	Criminal record check
4	Auto collision report
5	Uncategorized
6	Property liens
7	Environmental assessment, hazardous materials
8	Purchasing records, contracts
9	City emails, social media posts
10	Crime photo and video
11	Public works, utilities
12	Employee benefits, payroll
13	911/law enfo rcement service calls
14	Building code violations
15	City government meeting notes
16	Witness statements
17	Human services cases
18	Complaints to city
19	Checks and deposits

Analysis of Public Records Requests

https://sunlightfoundation.com/2018/10/16/results-from-analyzing-public-record-requests/

Most Popular Data Requests	
1	Police incident report
2	Parcel records, permits, plans
3	Criminal record check
4	Auto collision report
5	Uncategorized
6	Property liens
7	Environmental assessment, hazardous materials
8	Purchasing records, contracts
9	City emails, social media posts
10	Crime photo and video
11	Public works, utilities
12	Employee benefits, payroll
13	911/law enfo rcement service calls
14	Building code violations
15	City government meeting notes
16	Witness statements
17	Human services cases
18	Complaints to city
19	Checks and deposits

Analysis of WA State Department of Transportation (WSDOT)Public Records Requests

ODL Intern: Angela Gonzalez-Curci

https://medium.com/open-data-literacy/wsdot-records-as-open-data-731d525e0af3

Through this process, I found that the three most frequently requested record types: Crash Data, Right of Way Plans, and License Plate Readers, made up a plurality of all requests. Requests for the data and reports managed by the WSDOT Crash Data and Reporting Branch made up about 28% of the total.

Top Record Types	Count	% of Sample	Margin of Error
Crash Data	214	19.56%	0.09%
Right-of-Way	105	9.60%	0.07%
License Plate Readers	180	16.45%	0.08%
Total	499	45.61%	

Most Frequently Requested Record Types

Analysis of Public Records Requests

https://sunlightfoundation.com/2018/10/16/results-from-analyzing-public-record-requests/

- Consider proactively releasing datasets
- City of Seattle Open Data Risk Assessment
- “While cities can look to and learn from nationwide studies such as this one, there is still no substitute for targeted user research, to ‘ground-truth’ national trends and understand the local specificities of demand”

Sunlight Foundation

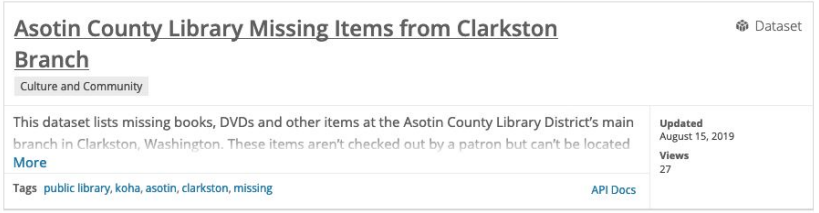
sunlightfoundation.com

- A Guide to Co-Design for Local Open Data Programs
  - “If open data programs hope to effectively deliver information to the public, they must be rooted in user needs. In addition to conducting user research, governments can help make this a reality by using co-design methods.”
  - Co-design: “participatory approach to designing solutions, in which community members are treated as equal collaborators in the design process.”
    - Key components of a co-design
      - Intentionally involving target users in designing solutions
      - Postponing design decisions until after gathering feedback
      - Synthesizing feedback from target users into insights
      - Developing solutions based on feedback
  - Develop use cases

Asotin County Library

- Intern: Lillian Curanzy, UW MLIS student
- Librarians: Jennifer Ashby, Mary Neuman

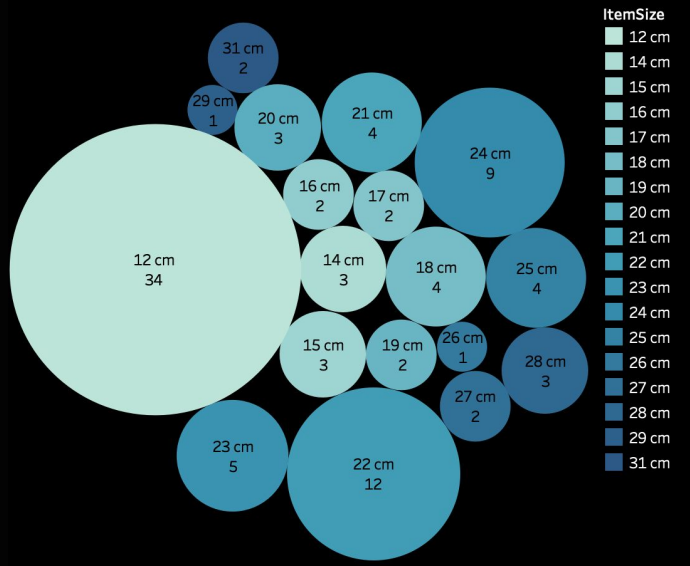
- ACL piloted the Data Equity for Mainstreet curriculum
- Hired Lillian through the UW Open Data Literacy project to advance civic open data
  - Created use cases based on interviews
  - Clarkston Fire Department annual call volume
- Published library data to the state data portal
  - https://medium.com/open-data-literacy/whats-not-there-c058a933e1bf



data.wa.gov/Culture-and-Community/Asotin-County-Library-Missing-Items-from-Clarkston/5de4-8ezn

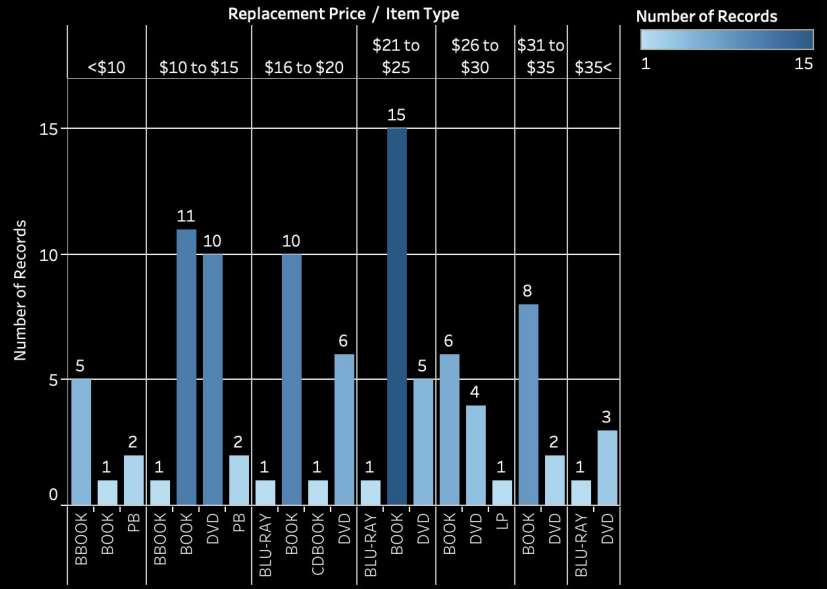
Missing Item Analysis

ItemSize



ItemSize and sum of Number of Records. Color lightens as item size decreases. Size shows sum of Number of Records. The circles are labeled by ItemSize and total number of items.

Item Price



Sum of Number of Records for each Item Type broken down by Replacement Price. Color shows sum of Number of Records. The marks are labeled by sum of Number of Records.

ODL
OPEN DATA LITERACY

Circulation by Item Analysis

Checkouts By Genre

Item Code (gro..	Call Number	Itype	
ADULT_NONFIC	270.092 Wes	NEWBOOK	71
	LD 248.843 Hol	LUCKYDAY	38
	O 741.5973 Wat v.3	BOOK	28
BIOGRAPHY	B Obama	BOOK	31
FICTION	Fic Bal	NEWBOOK	32
	Fic Ben	BOOK	31
	Fic Box	LUCKYDAY	48
	Fic Cob	NEWBOOK	37
	Fic Cou	LUCKYDAY	30
	Fic Dev	LUCKYDAY	30
	Fic Eva	LUCKYDAY	28
	Fic Gar	BOOK	34
	Fic Har	BOOK	40
	Fic Hil	LUCKYDAY	29
	Fic Jam	BOOK	27
	Fic Jan	NEWBOOK	36
	Fic Joh	BOOK	38
	Fic McC	BOOK	33
	Fic McM	BOOK	26
	Fic Mor	BOOK	63
	Fic Pat	BOOK	88
		LUCKYDAY	67
		NEWBOOK	28
	Fic Pen	BOOK	37
	Fic Rei	NEWBOOK	33
	Fic Rob	BOOK	72
	Fic San	NEWBOOK	38
	Fic Spa	BOOK	34
	Fic Ste	BOOK	28
		NEWBOOK	28
	Fic Tra	BOOK	30
	Fic Urr	BOOK	46
	Fic Win	BOOK	40
	Fic Woo	BOOK	64
	LD Chi	LUCKYDAY	45
	LD Cla	LUCKYDAY	27
	LD Eva	LUCKYDAY	45
	LD Fic Bur	LUCKYDAY	27
	LD Fic Pat	LUCKYDAY	69
	LD Fic Tod	LUCKYDAY	28
	LD Fic Woo	LUCKYDAY	33
	LD Mac	LUCKYDAY	27
	LD Ora	LUCKYDAY	27
	LD Owe	LUCKYDAY	40
	LD Pic	LUCKYDAY	33

Total Checkouts: 26 – 88

0 1 2 3 4

Number of Records

Sum of Number of Records for each Itype broken down by Item Code (group) and Call Number. Color shows sum of Total Checkouts. The marks are labeled by sum of Total Checkouts. The view is filtered on Item Code.

Popular Authors

= 20 checkouts

= 1 title by author

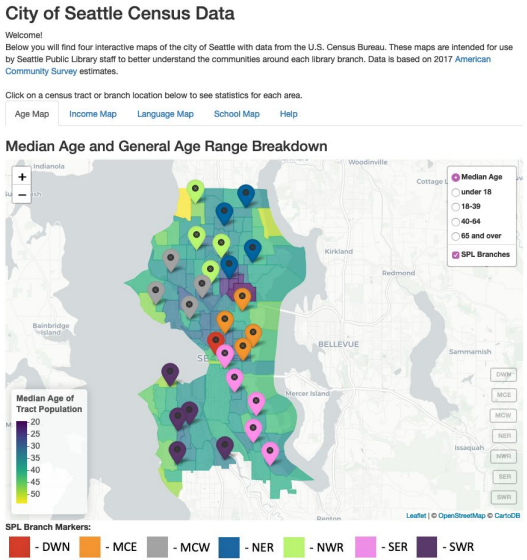
Total Checkouts: 26 – 250

Danielle Steel, C.J. Box, Luis Alberto, Lee Child, Jan Karon, Jodi Picoult, Debbie Macomber, Delia Owens, Don Winslow, Dav Pilkey, Rachel Hollis, Craig Johnson, John Sandford, Harlan Coben, Louise Penny, Judith A Jance, Liane Moriarty, Lisa Gardner, Nicholas Sparks, Elizabeth McCracken, Jodi Picoult, Taylor Jenkins Reid, David Baldacci, Marie Benedict, Michelle Obama, Catherine Coulter, Jude Devereau, P.J. Tracy, Anne Hillerman, Richard Paul Evans, Kate Morton, Charles Todd, Bill Watterson, James Lee Burke, Mary Higgins Clark, Marion Jensen, Debbie Macomber, Yusuke Murata, Tommy Orange, Angie Thomas, Kristian McMurris, Raina Telgemeier, Tara Westover, Nora Roberts, Stuart Woods, James Patterson

ODL OPEN DATA LITERACY

Seattle Public Library

- Intern: Karalyn Ostler, UW MLIS student
- Librarians: David Christensen, Francesca Wainwright, Wei Cai



- Using Open Data to Understand Communities Around Seattle Public Libraries
  - https://medium.com/open-data-literacy/using-open-data-to-understand-communities-around-seattle-public-libraries-fa3f9a96f3be
- Interviews with Regional Managers
- Create a Dashboard
  - Combines Internal Library Data with External Open Data
  - https://kostler.shinyapps.io/SPL-Seattle-Census-Data/
  - Code openly available: https://github.com/OpenDataLiteracy/SPL-KO

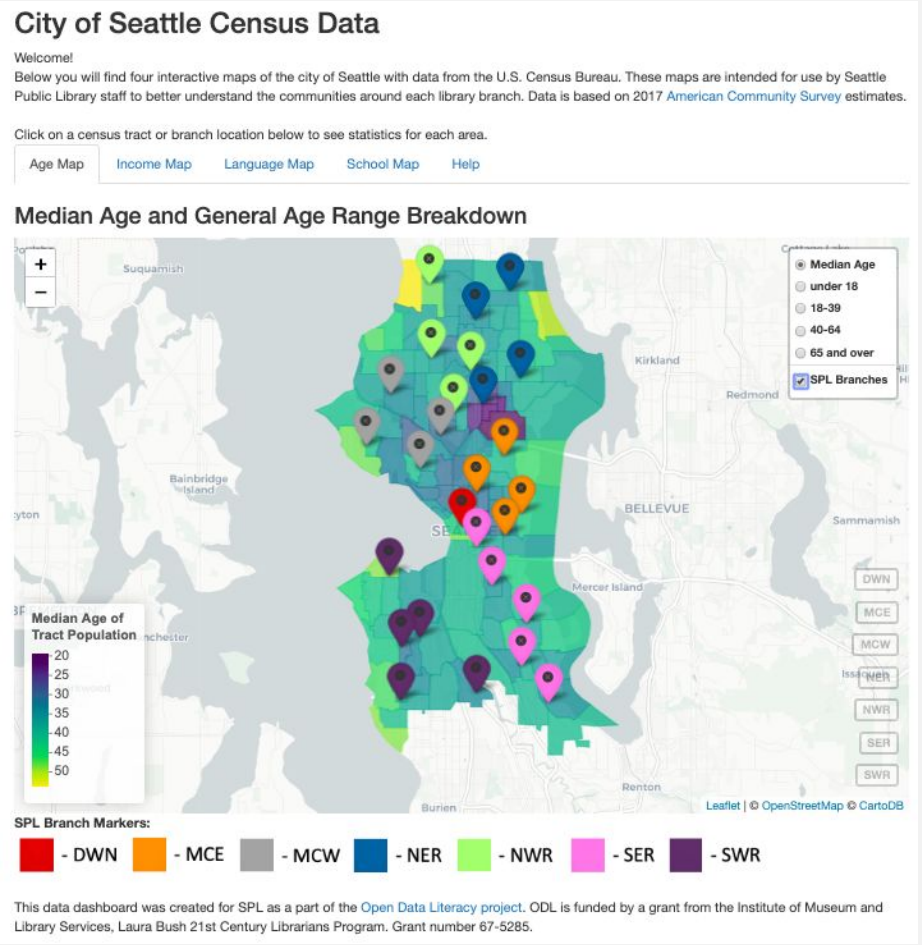
SEARCH FOR SUITABLE OPEN DATA

- Local Data:
  - City of Seattle Open Data portal
  - Washington State Open data portal

- Data.gov
- Packaged Data in R
- US Census Bureau

MAP DASHBOARD

- URL: https://kostler.shinyapps.io/SPL-Seattle-Census-Data/
- Browser based
- Hosted by RStudio
- Interactive
- Code easily modified:
  - Update with yearly census data releases
  - Add new data layers
  - Different city or library system



https://www.trl.org/open-data

My Branch | Contact Us | Donate | **My Account**

Timberland Regional LIBRARY

Books & More | Learn | Events | Locations | Kids & Teens | How Do I?

Home / About TRL / Open Data

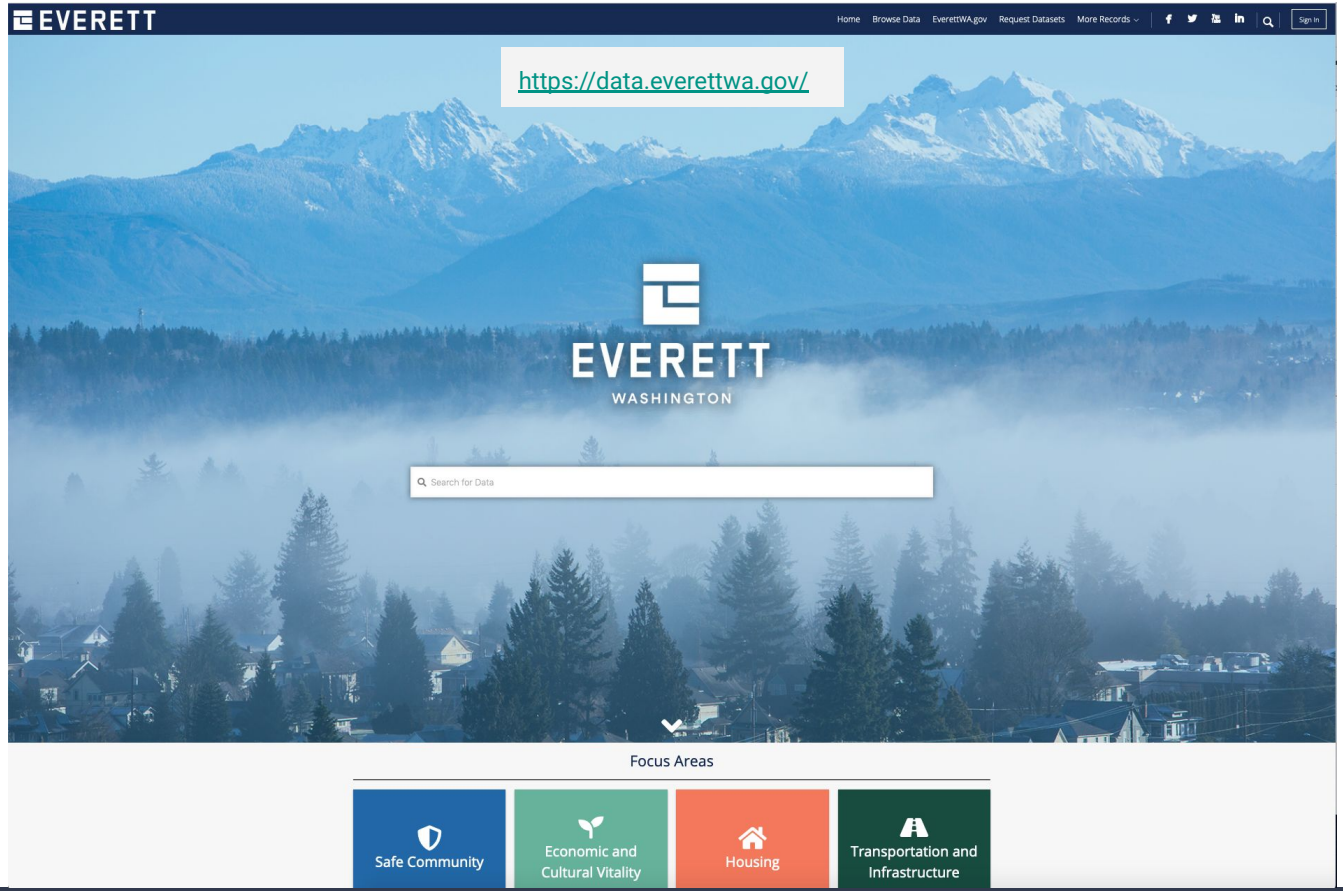
Search

	A	B	C	D	E
1	Branch	Year of Start Time	Month of Start Time	Sessions	Time Used (Min.)
2	AB	2017	02	756	26,729
3			03	3,256	122,757
4			04	3,102	110,978
5			05	2,986	110,362
6			06	3,126	112,605
7			07	2,879	106,082
8			08	3,245	120,180
9			09	2,958	116,779
10			10	3,317	130,642
11			11	3,019	119,858
12			12	3,008	119,582
13		2018	01	3,476	132,464
14			02	3,102	118,750
15			03	3,579	137,917
16			04	3,323	127,316
17			05	3,148	131,660
18			06	3,473	141,518
19			07	3,529	140,008
20			08	3,533	141,277
21			09	2,845	113,825
22			10	3,599	137,842
23			11	3,266	125,593
24			12	3,142	122,350
25		2019	01	3,523	137,505
26			02	2,916	114,086
27			03	3,399	132,012
28			04	3,627	146,394
29			05	3,204	123,678
30			06	1,969	81,745
31	AM	2017	02	19	316

Welcome to the Timberland Regional Library **Open Data Portal**

TRL recognizes the rights of citizens, organizations, and local government to public access to library records and information, as specified in the **Open Data Policy**.

Libraries and Open Data: Timberland Regional Library



EVERETT PUBLIC LIBRARY

Digital Equity

The City provides facilities to access and interact with our open data portal through the Everett Public Library. Internet computers are available to everyone in the community free of charge. Professional librarians are available to provide research and technological assistance.

Visit the Library ›

Libraries and Open Data: Everett

WA Public Libraries: Publish Your Data!

Publish some data!

Public libraries in Washington State can publish data on *data.wa.gov*, at no cost to the library and with **support from the State Library**.

Contact Kathleen to get started.

kathleen.sullivan@sos.wa.gov

About the Survey

- Surveyed public libraries, mostly systems but some individual branches, across Washington state
  - 51 respondents
  - 45% response rate
- Included questions about:
  - Current activities related to open data (e.g., digital collections, IT assistance, etc.)
  - **Interests, activities in open data**
  - Resources, capacity for open data
  - Alignment with mission, vision, etc.
  - Future actions

Thank you to all who participated in our statewide public library survey about open data!

And thank you to Washington State Library for their collaboration.

What data does your community need from the library?

Several Paths Emerge
in Survey Results

- Today, looking at results from one particular survey question:
  - **‘Considering the interests of your local user community, indicate your library’s likely or current level of activity, on a scale ranging from “Unlikely” to become active to “Highly active” now.’**
- Top two most active areas:
  - Data on Demand in the Library
    - Open Data at the Reference Desk
  - Open Data about the Library

43.14%

Active or **highly active** in

“reference assistance for community members on finding or using open data.”

33.34%

Active or **highly active** in

“publish[ing] our library’s data as open data.”

17.65%

Unlikely to

“build collections of open data.”

(second most unlikely activity)

Future Direction for Libraries

- Helping communities access the data they need
- Opening up more data, both at the library and beyond
- Moving beyond providing data, to curating data for the community

Resources

Open Data Literacy Project website: https://odl.ischool.uw.edu and blog posts: https://medium.com/open-data-literacy

Sullivan, Kathleen. (2018). 2018.12 Why Open Data Belongs at the Libraries. https://my.nicheacademy.com/washingtonstate/course/8739

Whatcom County Open Data: https://www.whatcomcounty.us/716/Data

Stern, Alena. (2018). What information do citizens want? Results from analyzing public record requests. https://sunlightfoundation.com/2018/10/16/results-from-analyzing-public-record-requests/

Gonzalez-Curci, Angela. (2018). WSDOT Records as Open Data. https://medium.com/open-data-literacy/wsdot-records-as-open-data-731d525e0af3

The City of Seattle. (2018). City of Seattle Open Data Risk Assessment. https://www.seattle.gov/Documents/Departments/SeattleIT/DigitalEngagement/OpenData/FPF-Open-Data-Risk-Assessment-for-City-of-Seattle.pdf

Jordan-Detamore, Greg. (2019). Introducing a guide to co-design for local open data programs https://sunlightfoundation.com/2019/08/22/introducing-a-guide-to-co-design-for-local-open-data-programs/

Data Equity for Mainstreet: https://data-equity.github.io/DataEquityForMainStreet/

Curanzy, Lillian. (2019). What's Not There. https://medium.com/open-data-literacy/whats-not-there-c058a933e1bf

Ostler, Karalyn. (2019). Using Open Data to Understand Communities Around Seattle Public Libraries. https://medium.com/open-data-literacy/using-open-data-to-understand-communities-around-seattle-public-libraries-fa3f9a96f3be

Timberland Regional Library Open Data: https://www.trl.org/open-data

Everett Open Data: https://data.everettwa.gov/

From Ambition to Action:
State & Public Libraries Partnering on Open Data

Cindy Aden, Washington State Library

Carole Palmer, Information School,
University of Washington

ODL

Information School
UNIVERSITY of WASHINGTON

SOS
Office of the Secretary of State
Washington State Library

- Background on open data & Open Data Literacy project
- Results from survey of Washington state public libraries
- Washington State Library experience and leadership
- Discussion

ODL

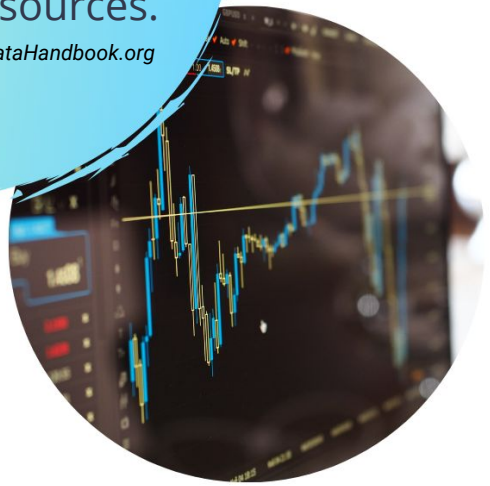
Information School
UNIVERSITY of WASHINGTON

SOS
Office of the Secretary of State
Washington State Library

What is open data?

Open data is data that can be freely used, re-used, and redistributed by anyone as publicly available resources.

Adapted from OpenDataHandbook.org



ODL

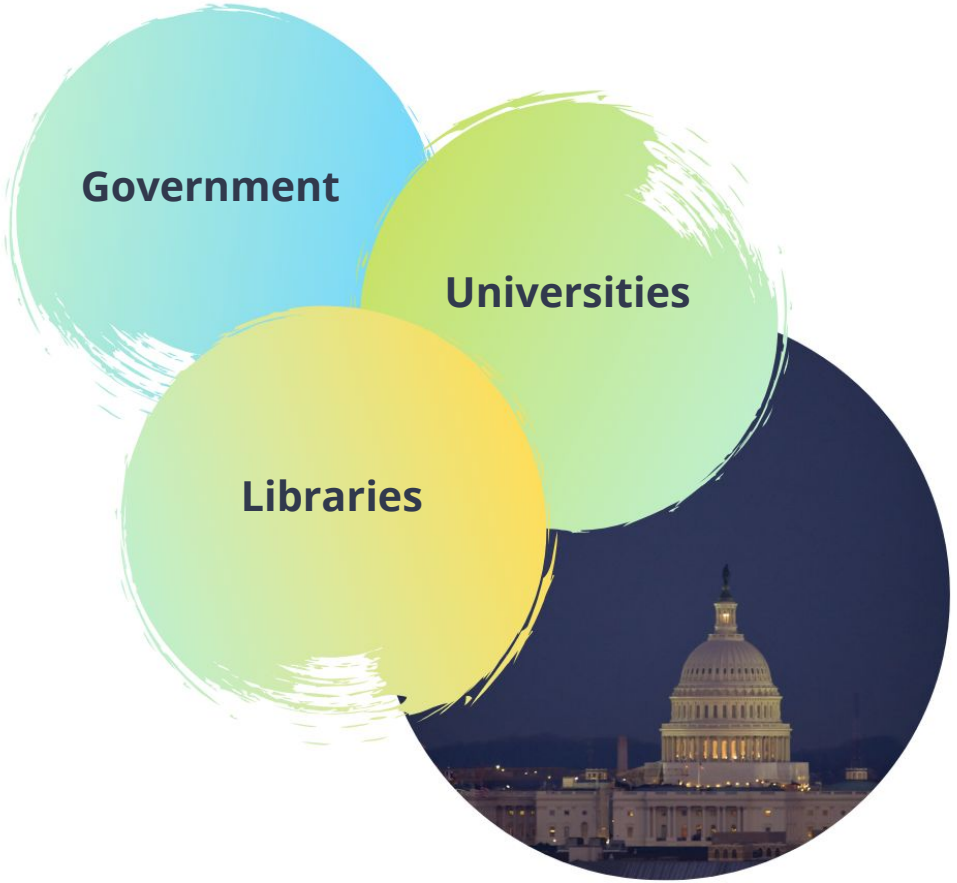
Information School
UNIVERSITY of WASHINGTON

SOS
Office of the Secretary of State
Washington State Library

Where is open data generated?

"Open data is powering a new civic movement that is changing the way citizens experience our nation's cities."

(Data.gov)



ODL

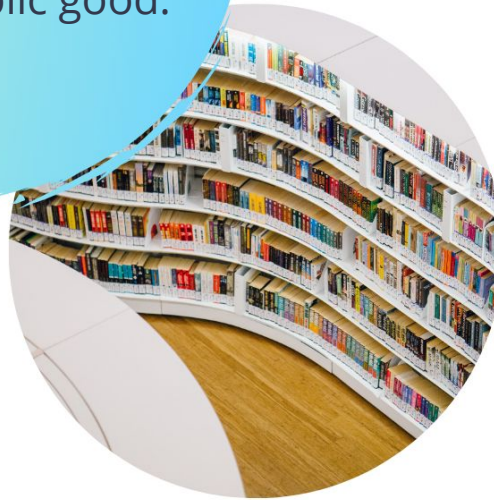
Information School
UNIVERSITY of WASHINGTON

SOS
Office of the Secretary of State
Washington State Library

Why should public libraries get involved?

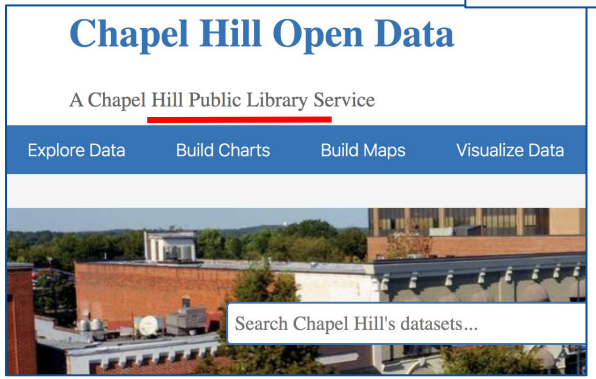
- Reference
- Collection development
- Information organization
- Information literacy
- Digital inclusion & ethics

Librarians have the professional expertise & community knowledge to make open data work for the public good.

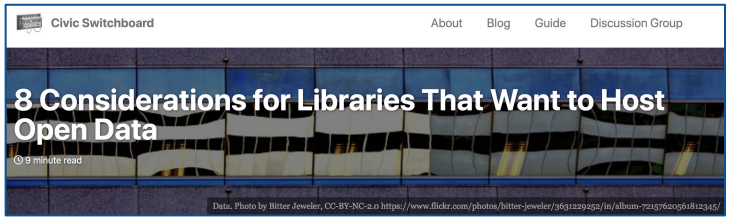


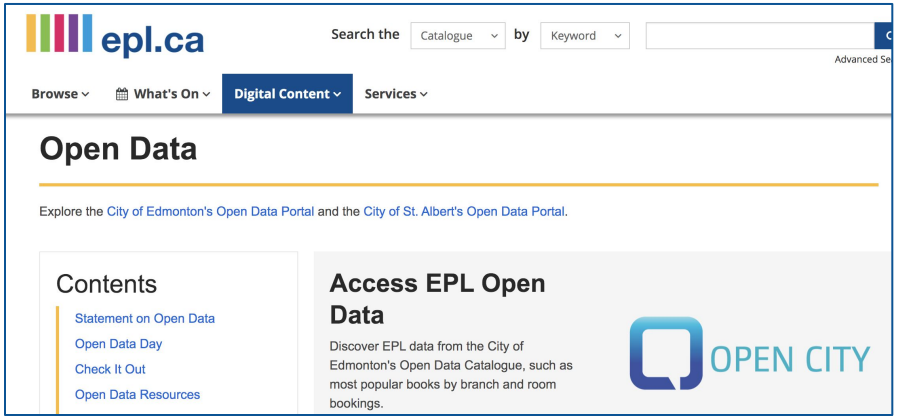
Chattanooga Public Library

NOT JUST FOR BOOKS: PUBLIC LIBRARIES PARTNERING WITH POLICE TO ENGAGE COMMUNITIES WITH OPEN DATA



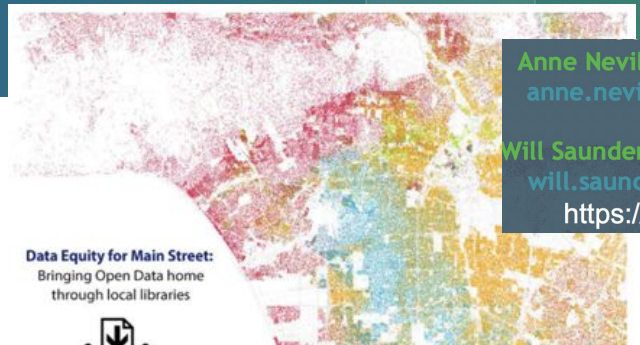
POLICE DATA INITIATIVE





DataEquityForMainStreet

A repository for curriculum generated by the Data Equity for Main Street project



Anne Neville / CA State Library
anne.neville@library.ca.gov *

Will Saunders / WA Tech Solutions
will.saunders@watech.wa.gov

https://data-equity.org

Open Data Literacy (ODL)

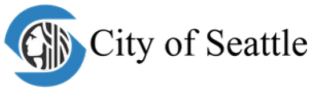
Open Data for Public Good - 2016-2020

Advance open data through education, research, and support for public librarians and information professionals in the public sector.

- New curriculum and educational resources
- Practical experience that benefit students, libraries, and government
- Partnerships with libraries and government agencies

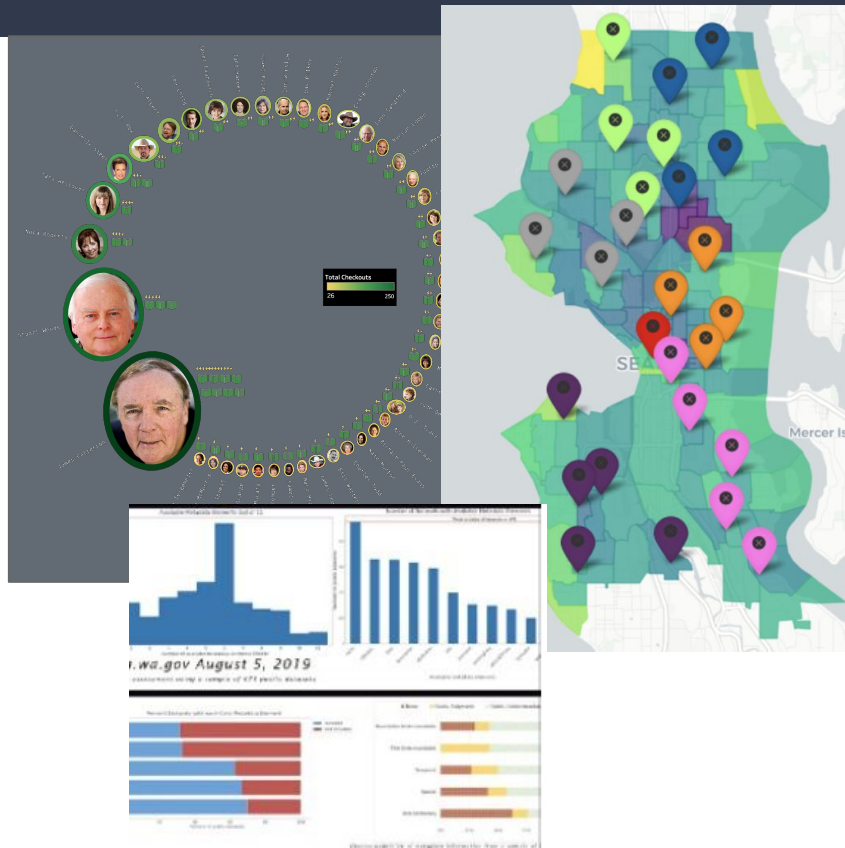
Internship partners

Placed 13 graduate students in 3 libraries & 3 government agencies



Selected 2019 internship projects

- Opening up library data in a rural library district
- Tool for applying census data to urban public library applications
- Quality assessment of state portal data for WSL



Public Libraries & Open Data Survey

Collaboration with Washington State Library to study interests, activities, capacity, and constraints in public libraries across Washington state.

Survey Design

- Current context
- Interest and activity in open data
- Capacity and resources
- Alignment
- Moving forward with open data

11 questions - 4 matrix, 6 open-ended, 1 multiple choice

Pretested from different professional perspectives

Survey Administration

- Awareness through WSL newsletter and social media
- **114 requests to participate**

  Via WSL email

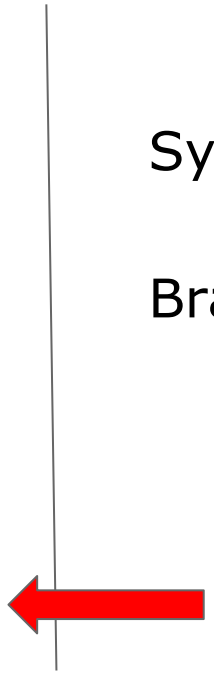
  - Two follow ups
  - Open for ~ two months

- 60 directors of systems
- 54 branch managers

  Representing 5 of 15 largest systems

Response rate - 45%

Under 5,000	9.80%	5
5,000 - 24,999	25.49%	13
25,000 - 99,999	15.69%	8
100,000 - 249,999	17.65%	9
Over 250,000	31.37%	16
Total	100%	51



Systems - 68.63%

Branches - 31.37%

Current Context and Alignment

> **90%**
provide

- guidance on information technologies
- guidance on new information resources
- information on community activities / resources

> **60%**
agree

open data initiatives align with

mission / community interests

43.14%

Active or highly active in

reference assistance

for community members finding or using open data.

22% somewhat active
16% likely to become active

Michelle Hawkins-Thiel. Sharing a Desk. Flickr

By service population:

- under 25,000 **44%**
- 25,000 - 99,999 **38%**
- 100,000+ **44%**

It's common to hear librarians lamenting the fact that they never get asked actual reference questions anymore. ...

But open data is an area where skilled guides are still needed. With the appropriate training, our librarians could be showing our community a richer way to view the world than through a simple google search.

I think that **given the opportunity, many librarians would jump at the chance to learn more.**

(-- Ben Haines, Senior Librarian at Sumner Pierce County Library, Service Population: 5,000 - 24,999)

33.34%

Active or highly active in

Central Library

	SPL	Census
Average Age	40.2	35.8
Under 18	4.0%	5.9%
18-39	52.1%	56.6%
40-64	34.5%	26.5%
Over 65	9.4%	11.0%

Publishing our library's data as open data.

- Under 25K - **44.4%**
- 25,000 - 99,000 - 37.5%
- 100,000+ - **24%**

Smaller libraries
(< 25,000 / n=18)

Most active:

- Collaborating with local government agencies to make their data available to the public
- Providing links to data portals

Twice as likely than large libraries to:

- Build collections of open data of interest to their communities

Overall

18% **Unlikely** to

build collections of open data of value to community.

20% **unlikely to offer programs** for community on benefits and use of open data.

Essential / high importance resources

- Increased operating budget 69%
- New or improved collaboration with city, state, or other partners 67%
- Professional development on access, use, and instruction 57%
- Support from data publishing platform 51%

Our rural library district has a very limited budget so we don't plan to pursue open data at this time / on our own.

Should we partner or gain support from grants, other libraries, or be approached by our local government entities, that could change.

(-- Rural Library, Service Population: 5,000 - 24,999)

Priority tensions

37% neutral, while 43% agree that

> Providing leadership for local open data initiatives should be priority for our library.

47% agree:

> It will be very difficult for our library to prioritize open data initiatives.

Kinds of support

• Online webinars	21%
• Regional in-person workshops	21%
• Online curriculum	16%
• Practicum students	11%
• Short-term consultant	11%
• Community volunteers	10%
• Paid interns	9%

Survey Takeaways

- Public libraries are well-positioned
- More than expected are already active, in a variety of ways.
- Education, leadership, and collaboration are key to progress

--technology and technical expertise not prominent in questions about professional development and staffing

Key Takeaways for

Washington State Library

Why support open data initiatives in libraries?

- Libraries are experts in data formation, maintenance, discovery and curation
- Libraries have right to play in facilitating better access to publicly available data

Implications for State Libraries

Can support open data initiatives by:

- Forging new government collaborations
  - WA Tech example
- Supporting professional development
- Creating just-in-time data applications for users and convening data creators
- Connecting libraries to funding opportunities

Leveraging ODL internships

Summer internships:

- 2018 - WA State Library and Data Equity for Mainstreet
  (Kathleen Sullivan)
- 2019 - Curation Assessment of Washington State Open Data Portal
  (Andrew McKenna-Foster)

New position, after 1st year:

  - Open Data Literacy Consultant

Progress after 1 year of consulting

- Supported 6 libraries in data collecting, publishing, teaching, and advising local government
- Opened up library data on budgets, circulation, programs
- Groundbreaking pilot partnership curating state open data
- Generated resource guide and other supporting materials

- *Future projects:* opening WSL data, data reference service

WSL

Open Data Literacy Consultant

Resource for the state library community

Kathleen Sullivan

kathleen.sullivan@sos.wa.gov

360-570-5577

Next Steps - http://odl.ischool.uw.edu/statelibrary-nextsteps/

Getting Started

- WSL Open Data Guide
- ODL Resource Guide
- Data Equity for Main Street resources
- WSL / ODL webinars

Survey

Extending beyond WA with support from Technology and Social Change Group (TASCHA)

- Contact Cindy or Carole (clpalmer@uw.edu)
- Email ODL at odl@uw.edu

Discussion

How similar or different are the opportunities and challenges in your state?

What could be gained from a broader survey?

--value for state libraries, COSLA, profession?

THANK YOU

http://odl.ischool.uw.edu/statelibrary-nextsteps/

ODL

Information School
UNIVERSITY of WASHINGTON

SOS
Office of the Secretary of State
Washington State Library

An Educational Workbench for Data Curation

Carole L. Palmer and Tom Johnson

Information School
University of Washington

7th NDS Consortium Workshop
Chicago, IL
April 14, 2017

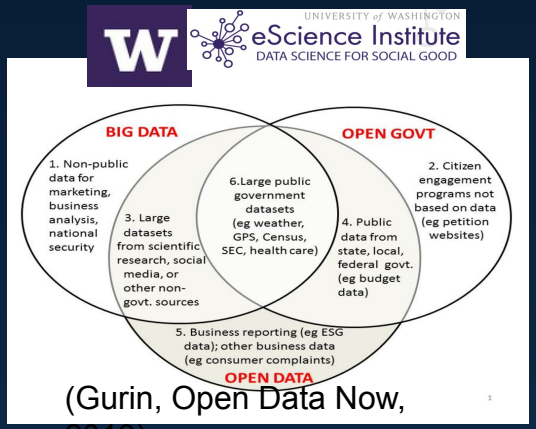
UW iSchool Educational Context

Data curation signature within iSchool MLIS – research data focus with additional Data Science sequence in MSIM

Expanding curriculum

- data services and technology
- government and civic open data

Extending DCERC field experience model to public sector – making open data more usable



(Gurin, Open Data Now, 2013)

democratizing data
entrepreneurial use

DCERC

NCAR

ODL
OPEN DATA LITERACY

Institute of Museum and Library Services

Department of Ecology State of Washington

Washington Department of Fish & Wildlife

Washington State Historical Society

Washington State Department of Transportation

The Seattle Public Library

Washington State Office of the Chief Information Officer

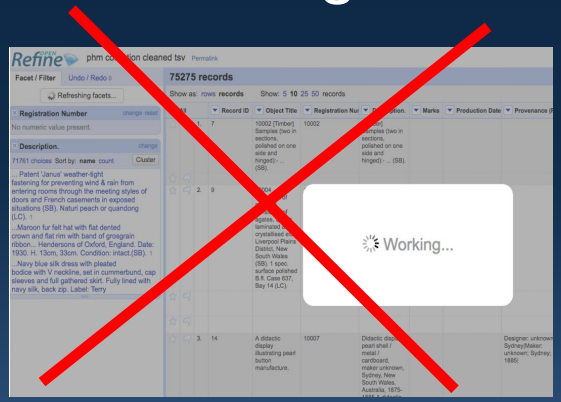
Seattle Information Technology

Educational Workbench Overview

- Idea spawned in breakout at 6th NDS workshop
- NDS Labs Workbench platform

  better prepare students for real world workforce demands

  tools for data curation & data services customized for practical, experiential online learning



- Reduce barrier to access and use
  - simple sign on
  - on demand service launch
  - integrate multiple platforms and tools into coursework

Objectives

Support hands-on experimentation for 30-50 students annually

- build purposeful data collections
- data packages for ingest, archiving, dissemination
- perform harvesting, manipulation, indexing, metadata normalization & enhancement
- informed examination of existing platforms & services

Support range of pedagogical approaches

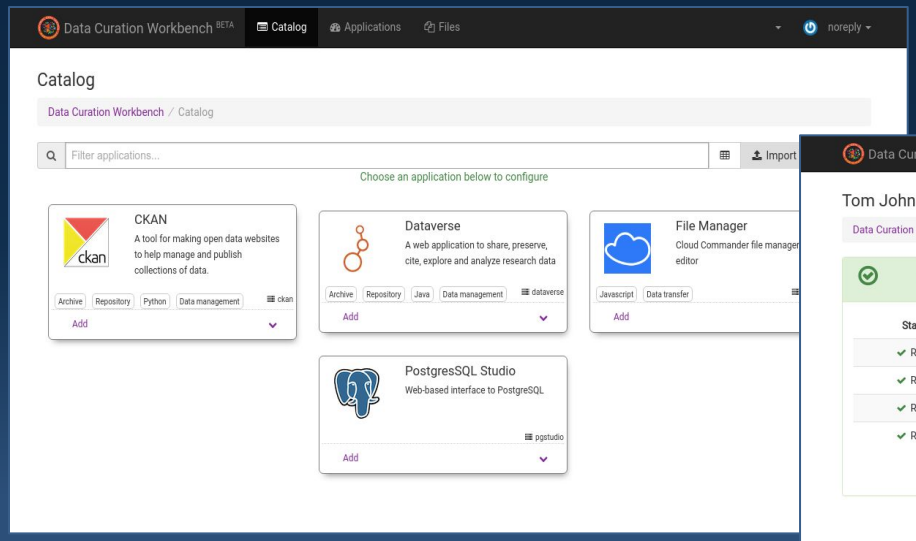
- structured individual & group assignments
- independent exploration
- sharing of expertise, group problem solving
- peer and instructor review & exchange

First steps

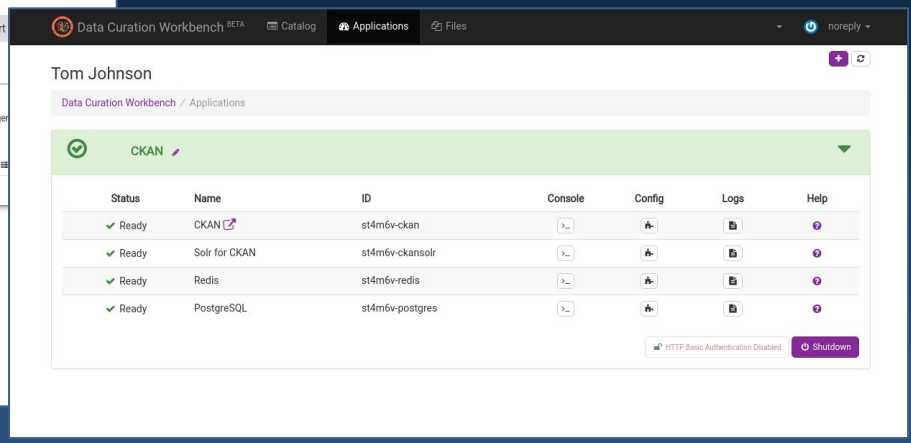
Spring 2017 - Dataverse and CKAN core

Advanced Data Curation themes – data quality, interoperability, granularity, complex objects, versioning, catalogs & collections

- Options to implement at no risk to grades
- Assignment to identify and evaluate candidate tools



Workbench with initial 4 applications.



Launched CKAN instance

Timeline

Summer - Fall 2017 development

- suite of tools across curation workflow
- pedagogical data sets and assignments

Winter 2018 launch in Fundamentals of Data Curation

Labs to complement content areas, including:

levels of service; reproducibility, transparency, & reuse; metadata for discovery/use; identifiers; access & licenses; sensitive data & privacy; citation; preservation

Next Steps

- Better integration with courseware
- Expand the application catalog

  Support a broader range of curation tools
  Add programming environments to support API interactions

- Develop assignments and self-guided materials for courses
- Incorporate student feedback

Tentative Future Plans

- Pursue funding to extend to support broader user community

  Other courses - Digital Preservation, Databases, Data Science

  Other iSchools

  Continuing education for professional communities
    - Workshops
    - Independent learning

- Shared online educational materials

Thank you

Carole Palmer
clpalmer@uw.edu

https://www.youtube.com/watch?v=N2zK3

Preparing Masters Students for Public Sector Data Service

Nic Weber
University of Washington - Information School

November 12
Vancouver, BC - ASIS&T 2019

Information School
UNIVERSITY of WASHINGTON

The public sector is increasingly engaged in the management, publication, and analysis of structured data.

iSchools should be well positioned to produce students that can help:

- **Citizens** realize the potential of increased access to open data
- **Governments** in both curating and publishing open data
- **Non-profit organizations** combine and analyze different open data sources

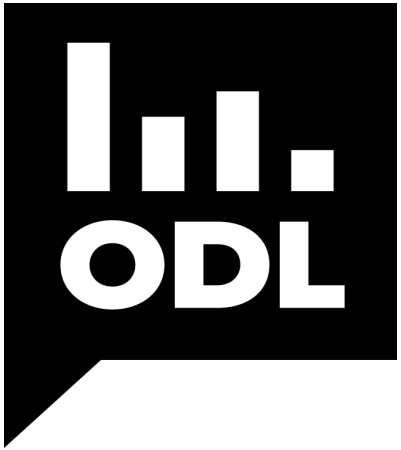
But....

- **Data science** programs are often aimed at developing analytic expertise that is highly sought after (and well compensated) by **industry**

- **Data curation** programs are aimed at developing a workforce that is general in skills, but most effective in specific **scientific domain** applications.

Open Data Literacy (ODL)

- Institute of Museum and Library Services (IMLS) funded 4 year project.
- University of Washington iSchool
- Local Partners:
  - Seattle Public Library
  - WA State Department of Transportation
  - City of Seattle
  - WA CIO's office
  - WA State Library



ODL Goals

- **Curriculum**: Formal and informal learning modules using open data from public sector - audience is both iSchool students, and current LIS professionals
- **Workforce:** Build capacity for Information Professionals in the Public Sector to curate and improve the accessibility of open data
- **Partnerships:** Develop a network of sustainable partnerships with open data stakeholders
    - **Guest Lectures:** Bring real world experience and examples into classroom
    - **Capstones + Internships:** Mutually beneficial learning opportunities for students and partners
    - **Community Building:** Create extended learning opportunities for working information professionals and communities of interest

Successes (so far)

- Internships
  - 8 x 8 week funded opportunities. Sponsors provided actionable ways to improve public records management, metadata reconciliation, and outreach planning.
- Capstones
  - 9 public sector sponsors for over thirty students.
  - “This team outperformed most of the contractors we’ve hired to work on this map”
- Curriculum
  - **Foundations:** Database Concepts for Information Professionals; Introduction to Programming for Information and Data Science; Introduction to Metadata Design
  - **Intermediate**: Data Curation 1 + 2; Interactive Data Visualization; Introduction to Data Science
  - **Advanced:** Data Science 1, 2, 3; Social Media Data Mining; Advanced Relational Databases

Challenges (so far)

1. Public sector domain knowledge
    a. Citizens ! = Experts in local government
2. Sequencing data-based coursework for success in public sector
    a. “I know that geoJSON is the right choice for structuring this kind of data... I just didn’t know how to convince my project sponsor it was the right choice”
3. Scaling and Sustaining proof of concepts
    a. Students often want to develop requirements, build, and evaluate... Rarely do sponsors get beyond the ‘develop requirements’ stage
    b. Proof of Concepts are ways to demonstrate potential impact, but sponsors are left with knowledge THAT open data programs can be improved with little capacity to scale or sustain

My Question for You...

- **Open Data in Classroom**: How do you use open data, or public sector information in your curriculum and coursework? What unique challenges does open government data pose for your data science and data curation curriculum development ?

- **Scale**: Many practical applications of data science and data curation require students to not just understand how to solve a problem, but how to scale a solution... How does your institution address this issue in internships, capstones, coursework?

Open Data Literacy Retrospective

Carole Palmer

ODL Project Reunion
June 5th, 2020

Classroom preparation

Fundamentals of Data Curation / Data Curation 1

117 students

Independent studies

Capstones

Research projects

Advanced Data Curation / Data Curation 2

40 students

Digital Preservation

32 students

14 Internships

2017-2018

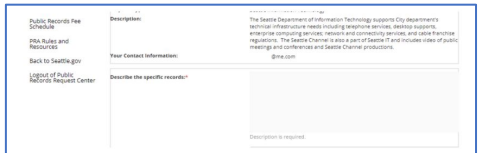
Can Libraries Accelerate Local Open Data Publishing?

Kathleen Sullivan

WSDOT Records as Open Data

Angela Gonzalez-Curci

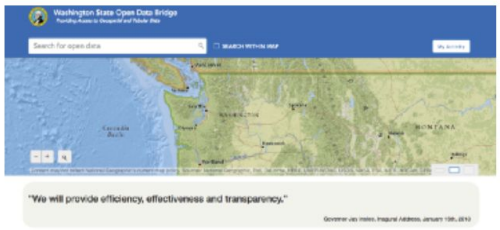
Kevin McCraney



Improving User Experience for Public Records Requests

Ruminations and meditations on public records...

M. Wynn Tranfield



OCIO Geospatial Program & Open Geospatial Data

Leslie Denning

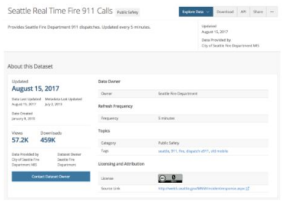
Nina P. Showell

Tim Blankemeyer

Sarah Carrier

Wrapping up with Public Records & the Open Data Alliance in Washington

Using Metadata Standards at the City of Seattle

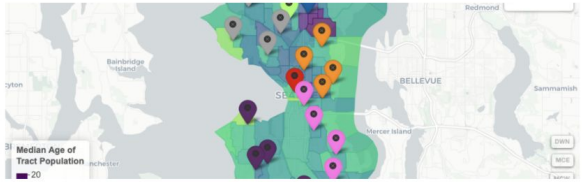


ODL project builds foundations for future of innovative WSDOT data catalog

Engaging with Open Data: Seattle Public Library's Role in Open Data

2019-2020

congratulations graduates



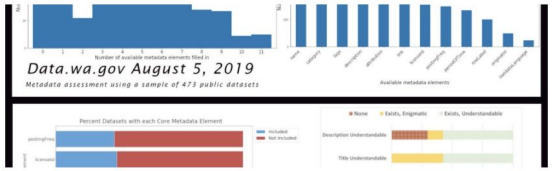
Using Open Data to Understand Communities Around Seattle Public Libraries

Karalyn Ostler

Opening Rural Library Data: Applying Coursework to the Field

Andrew Mckenna-Foster

Lillian Curanzy



Facing the Flood: Assessing Metadata Quality on Washington's Open Data Portal

Joan Hua

Laying the Groundwork for Transportation Research Data Management

Kaitlin Throgmorton

Jamie Ramos

The Myth of the Obvious

Is there such a thing as 'too obvious' metadata? This article explores why we...

Tracking Washington State Public Library Responses in Crisis

Continuing activities

Open educational resources

Student preparation through established curriculum

With WSL Open Data and Public Libraries survey, extending to other states

– supported by Technology and Social Change Group (TASCHA)

No cost extension on IMLS grant through March 2021

Dissemination
Some selected student field experiences

Strategize on next grant to advance activities and impact

<u>What activities are of most interest to our partners?</u>

Applause

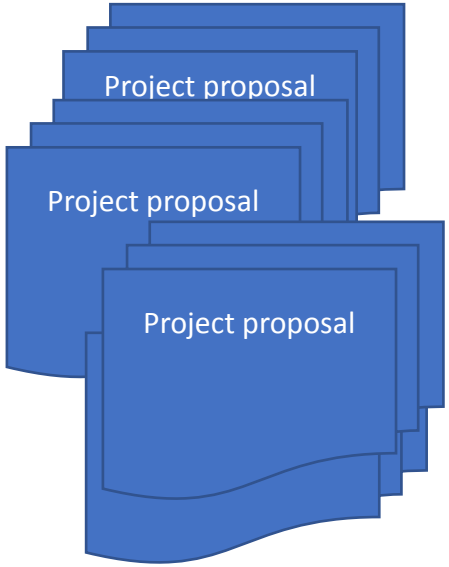
For your all your incredible contributions and dedication

Applause

ODLI

10 Recommendations for Creating Internships with the Public Sector

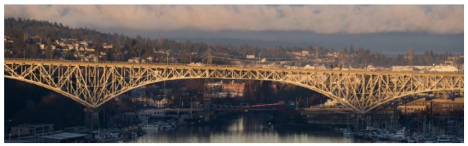
Bree Norlander



Kaitlin Throgmorton

Public Libraries & Open Data Survey: Initial Findings

Open data: New information resources for the library's enduring mission



What makes open data good, and how can we make it better?

WA Office of the Chief Information Officer

Will Saunders

City of Seattle

David Doyle
Paul Alley
Subu Swayam
Synba Wang
Claire Palay

WA State Library

Cindy Aden
Kathleen Sullivan
Evelyn Lindberg

The Seattle Public Library

Jim Loter
David Christensen

WSDOT
Washington State Department of Transportation

Leni Oman
Gordon Kennedy
Andy Everett
Kathy Szolomayer

Asotin County Library

Mary Neuman
Jennifer Ashby

Looking forward to future interactions.

Information School
UNIVERSITY of WASHINGTON

Appendix 5:

Representative flyers & handouts

Interested in Open Data?

ODL
OPEN DATA LITERACY

What is open data? Open data is data that can be freely used, re-used, and redistributed by anyone as publicly available resources that can be accessed, used, and shared by all. (Open Knowledge International, Open Data Handbook.) Open data promotes transparency, citizen involvement, innovation, informed decision-making and action, to the benefit of individuals, communities, government, and business.

What is Open Data Literacy (ODL)? The ODL project recognizes the importance of the open data movement for public institutions, especially libraries and government agencies. Our work responds to the need for librarians and information professionals to develop expertise to advance open data initiatives in the public sector and to foster open data literacy among citizens and communities. To prepare students for professional roles in open data, ODL is developing curriculum and providing fieldwork experiences, in partnership with government agencies and public libraries, as we also study the evolving trends, challenges, and workforce demands in open data.

The following iSchool courses are recommended for students seeking preparation in open data, and for students interested in competing for ODL summer internships in libraries and government agencies.

RECOMMENDED COURSEWORK

These courses have been designed by the ODL team to integrate open data knowledge and expertise:

LIS 545: Fundamentals of Data Curation*
LIS 598: Digital Preservation*
LIS 598: Data Curation 2
LIS/IMT 511: Intro to Programming
IMT 562: Interactive Information Visualization
LIS/IMT 572: Introduction to Data Science
IMT 573: Data Science I*
IMT 574: Data Science II*
IMT 575: Data Science III

*Winter Quarter Offerings!

2019 SUMMER INTERNSHIPS

Internship opportunities will be available for students to apply their data curation and data science knowledge and expertise to real-world open data projects. ODL expects to place four students in paid internships in libraries and government agencies, such as the Washington State Library, City of Seattle, Washington State Department of Transportation, and Washington State Office of the Chief Information Officer. The positions will be advertised through iCareers during the 2019 Spring Quarter.

@ODLiteracy

odluw@uw.edu

RECOMMENDED COURSES

INFX 551 A: Fundamentals of Data Curation

Description: Data curation is an emerging area of expertise for information professionals across sectors. This course examines the curation of digital research data for professionals working in institutions involved in data intensive research, scholarship, and innovation. The course covers principles and practices in data management and data services, and emphasizes curation expertise for development of reusable and accessible data resources, within the context of current trends in data sharing and open access.

Example Topics: Institutional and Disciplinary Perspectives on Data Curation; Metadata & Provenance; Data Services; Architecture and Preservation; Attribution and Reproducibility; Sustainability and Costs

INFX 598 D: Advanced Data Curation

Description: This course examines a broad range of practical and conceptual issues in the emerging field of data curation. It focuses on recent advances and challenging problems in the curation of research data across disciplines and new trends in open data resources and services for research communities and the public. The course will draw on current and foundational literature and research, as well as current initiatives, to examine key challenges in the field and practical solutions applied by data professionals. In-depth study of topics will be driven by student interests and their contributions to the course through discussion, current awareness briefings, and their final project.

Example Topics: Data Curation Evaluation; Cultures of Data/Evidence; Inventories, Catalogs, & Interoperability; Data Levels, Provenance, Process; Value & Quality; Enhanced Publications and Complex Research Objects; Physical Samples and Data Rescue; Open Data & Public Libraries

INFX 598 OD – Open Data Infrastructures and Services (under development)

Description: This course will prepare students to design, implement, and sustain open data initiatives. This will be achieved through a combination of lectures and laboratory work. Lectures will be based on case-studies from institutions that have engaged in open data publishing, and address topics such as privacy, transparency, interoperability, trustworthiness, stewardship, and costing & sustainability.

Hands-on laboratory instruction will include two components:

1. Techniques for retrieving, cleaning, and visualizing open data.
2. Designing and deploying open data portals through the use of CKAN software.

The course will be designed around a single group project which will require students to launch an open data portal using the open-source software platform CKAN. This exercise will require students to gather requirements from a potential user community, locate necessary data and metadata necessary to serve that community, and then design policies for publishing and providing access to these collections.

Example topics: Best Practices for Publishing Open Data; Retrieving and Cleaning Open Data; Privacy, Politics, and Economics of Open Data; Describing and Creating typed relationships between Open Datasets; Preservation, and Sustainability of Open Data Initiatives.

Related recommended courses (not required)

INFX 572 Introduction to Data Science (4)
Introduces a broad, non-technical overview of key concepts, skills, and technologies used in "data science". Provides a high-level introduction to common data science pipelines, such as experimental design, data collection and storage, basic analytics, machine learning, and data visualization, focusing on analyzing in real-world datasets using industry standard statistical packages.

INFX 573 Data Science I: Theoretical Foundations (4)
Provides an overview of key concepts, skills, and technologies used by data scientists, including inference; machine learning and pattern recognition; storage and scaling; experimental design; and data visualization.

INFX 574 Data Science II: Machine Learning and Econometrics (4)
Provides skills required to analyze and derive insight from large-scale, heterogeneous data. Covers key concepts of functional and imperative programming for storing, extracting, analyzing, and presenting large data projects; and data analysis using inferential statistics, supervised and unsupervised machine learning. Students gain experience modeling social and behavioral data.

INFX 575 Data Science III: Scaling, Applications, and Ethics (4)
Advanced skills for analyzing and deriving insight from large-scale, heterogeneous data. Provides methods, tools, and frameworks for analyzing data, with topics including scaling and distributed computing, network analysis, and visualization. Student gain experience applying methods to real-world data.

INFX 598 – P Digital Preservation

Description: Libraries, archives, and museums are increasingly charged with the curation and long-term preservation of digital objects – including collections of cultural heritage objects, born-digital scientific data, new digital media (such as video games), and software. This course focuses on tools, techniques, and policies for delivering trustworthy preservation in these settings. Practical activities will include the design and implementation of a digital preservation plan, using established frameworks to audit preservation services, and learning digital preservation standards such as reference models, file formats, metadata schemas, and certifications for trustworthy digital repositories.
Example Topics: a basic introduction to preservation strategies, such as migration and emulation of digital resources; preservation standards and specifications for providing long-term access to digital material; forensic approaches to profiling complex digital objects; sustainability and risk assessment, including digital repository auditing frameworks and certifications; as well as the extension of traditional archival concepts like authenticity, integrity, quality control, and provenance to a digital realm.

Last Update: Oct 27th, An Yan

Information School
UNIVERSITY of WASHINGTON

Open Data Literacy (ODL) is a collaborative project led by the Information School (iSchool) at the University of Washington (UW) in partnership with public institutions that create, manage, and publish open data. Current partners include the Seattle Public Library, Washington State Historical Society, Washington State Department of Transportation, and the Washington State Office of the Chief Information Officer.

ODL
OPEN DATA LITERACY

The overarching goal of ODL is to improve public accessibility and use of open data by increasing the data capabilities of both data producers and data consumers. To meet this goal, ODL will advance data literacy through new curriculum for iSchool students, internships and fieldwork at institutions with open data initiatives, and community workshops and open education resources for public librarians and information professionals.

Curriculum: ODL curriculum will prepare students and practicing information professionals to:

- Curate collections of open data of value to local communities
- Build infrastructures and preservation environments to sustain open data collections,
- Collaborate with open data providers on advocacy and outreach activities.

Student Internships and Fieldwork: UW iSchool students with data curation and data science expertise will be placed in partner organizations, to apply their classroom knowledge to real-world open data projects. Placement options include paid summer internships for intensive 8-week projects and directed fieldwork placements for longer 2-6 month projects. Areas of focus for ODL student projects include:

Data Collections on Civic Concerns	Professional Best Practices
• the environment	• open access and use
• natural resources and hazards	• interoperability
• regional cultural heritage	• digital preservation

Community Workshops: In collaboration with project partners and UW personnel, ODL will host open data workshops in 2017 and 2018. These public events will engage particular communities in using open data through hands-on instruction and small group activities.
Open Education Resources: Over the course of the project, ODL will offer continuing education opportunities for information professionals to learn how to manage, curate, or preserve open data through webinars, an instructional blog, and use of open-source software.

To find out more about ODL, please visit:
http://odl.ischool.uw.edu
For information on institutional partnerships and education and outreach activities, contact:
odluw@uw.edu.

Version 1, last modified: Oct 27st, 2016

OPEN DATA LITERACY

STUDENT PROJECT DESCRIPTION TEMPLATE

Date:

Organization:
Name of Contact:
Contact Email:

1. **General description of the project (100 words or less)**

2. **What is the topic of theme of the project? ***

   - ☐ *Natural resources, hazards, or public safety*
   - ☐ *Cultural Heritage*
   - ☐ *Community Engagement*
   - ☐ *Others*

     Please specify:

3. **Has there been any previous work on this project**

4. **Is there any specific software or data that will be used for this project**

5. **How long will the project take to complete (estimate in hours):**

1

6. **Is there a deadline for the project to be completed?**

7. **What are the (anticipated) deliverables?**

8. **Who will be the mentor for this project (name and title):**

9. **How many students could the project support?**

10. **What skills will be necessary to complete the project (e.g. Knowledge of SQL databases; Data Management Planning, etc.)?**

11. **What skills will be learned during the project?**

12. **Does your organization have funding to support this project?**

13. **Any conditions or restrictions that may be applicable (team size, restrictions on public display of institution or project info, work authorization restrictions, intellectual property, etc)?**

2

Information School
UNIVERSITY of WASHINGTON

OPEN DATA LITERACY

STUDENT FIELDWORK OPTIONS

A major part of our grant work will be to broker field experiences between you as project partners and students at the UW iSchool.

1. Capstone Projects

"Students serve as consultants to an organization and identify an information problem or need. They then develop the approach and methods needed to address the problem, conduct the research and present the findings in both oral and written formats. Capstone projects are developed during winter quarter (January-March) and are implemented during spring quarter (April through June)." (For more info see: https://ischool.uw.edu/capstone)

Capstone projects are posted on iCareers. Students are required to establish their projects by Dec. 15, 2016. Sponsor Networking Events are held in October, and students are looking for projects in Oct. through Dec. on iCareers. Sponsor's website https://ischool.uw.edu/capstone/sponsorship .Official Capstone guide can be found in : https://www.dropbox.com/s/pc8pnq1mmuxuv00/2016%20Capstone%20Proposal%20Guide.pdf?dl=0

For ODL records, we need a completed project proposal form for funded or unfunded work (we very much welcome feedback on this form): https://docs.google.com/a/uw.edu/forms/d/12_MjkvtCSDZmmik4F5b68kfErAWRvFZFX920NI4Vsz4/edit

2. Directed Fieldwork

"Students work under the formal guidance from a qualified host site supervisor (mentor) during their DFW experience. Qualified host site supervisors have an MLIS, MLS, MSIM, library media endorsement (for teacher librarians), or other relevant degree as well as the depth of experience and strong professional identity of an information professional. Also, the supervisor should be more knowledgeable than the student for most of the DFW tasks." (See: https://ischool.uw.edu/current/mlis/fieldwork-internships/host-site-info)

Students may pursue fieldwork according to these credits/hours:

- 2 credits = 100 hours
- 3 credits = 150 hours
- 4 credits = 200 hours
- 5 credits = 250 hours

Supervisor responsibilities include:

- Providing appropriate supervision and guidance, including regular (weekly) meetings between the supervisor and student to discuss activities, monitor progress, and discuss any challenges of problems
- Providing relevant tasks, projects and activities that constitute beginning professional level work and involve skills and knowledge that the student has already begun to develop through coursework but for which he/she has not yet developed a high level of expertise
- Working with the student on the following paperwork:
  - The initial Learning Outcomes Agreement
  - The Midterm Report
  - The DFW Supervisor's Final Evaluation

3. Internships

ODL internship will be a new, competitive field experience option for summer placements. Students will work with data mentors at their field site on projects designed to directly benefit partner organizations and local service communities. Paid interns will have additional responsibilities to work with their mentors and ODL team to create professional development resources to be disseminated and used by practicing librarians (e.g. case studies, technical documentation, and best practices guidelines). ODL funding will cover:

- 5 students in Years 1 & 2 - 10 total
- 8 Weeks (40 hours week)