Access, Accountability, and Ownership in Government Use of Proprietary Systems

Meg Young

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Megan Finn, Chair

Ryan Calo

Nic Weber

Jan Whittington

Program Authorized to Offer Degree:

Information School

University of Washington

Abstract

Access, Accountability, and Ownership in Government Use of Proprietary Systems

Meg Young

Chair of the Supervisory Committee:
Megan Finn, Assistant Professor
Information School

When firms contract with public agencies to provide services, they regularly assert that some subset of their work is proprietary. For instance, companies may stake a claim over how information they produce is managed and shared. At the same time, governments in the state of Washington are subject to a strongly transparent state Public Records Act, under which members of the public can request access to government information, some of which is transacted or shared between government agencies and private firms. In this dissertation, I analyze two cases of contracting relationships between the public and private sectors in which data ownership is contested: (i) the contract between the transportation agencies behind One Regional Card for All (ORCA) fare card in the Puget Sound region and their vendor Vix Technology and (ii) the would-

be contract between King County Metro and Lyft Incorporated in support of a subsidized expansion of transit hub access. I report on data sharing as a site of competing claims over data control with a focus on the technical, legal, and policy factors that enabled and constrained access to it. In both cases, I situate specific dataset access requests in the context of public agencies' objectives to advance accountability, equity, and oversight. Meanwhile, firms asserted intellectual property protections like trade secret in an attempt to prevent data access. My data collection draws on my firsthand experience as a research assistant on a project to support cross-sector data sharing at the University of Washington called the Transportation Data Collaborative, as well as additional interviews I conducted and documents I requested under the Washington Public Records Act. In its high-transparency and permissive data access construction, the Washington Public Records Act emerged as a key factor in both of my case studies, as it both afforded and constrained data sharing in moments when firms contested access on the basis of trade secret. I locate my descriptive case studies of these data flows within a broader scholarship of ambiguity and political struggle in the demarcation between "public" and "private." Building on theory that understands these terms as always making normative (rather than descriptive) claims about the world, I explore how ownership emerged as a dominant rationale in assertions about how data should be accessed or shared. Across both cases, I observe that while data ownership is locally understood to be a means of asserting data control, in practice it is not dispositive of outcomes with respect to its access and sharing. In the conclusion, I highlight how data ownership fails as a data governance strategy and connect this finding to broader information policy debates.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

So many people made it possible for me to carry out this work. In particular I would like to thank my advisor, Megan Finn, whose staggering command of vast areas of literature was invaluable as I looked for ways to orient this project. She helped me to untangle and weave back together the threads I was left with after years in the field. Beyond her role in strengthening and sharpening this dissertation, my discussions with Megan on theory, method, domain, research ethics, and practice helped me to develop as a scholar. Her dedicated feedback consistently supported me to find and inhabit my own voice, positionality, and ideas.

I am so grateful to have worked with Jan Whittington and for her generous mentorship over time. Through working together, I learned how to step outside a traditional researcher role to work directly with public agencies and firms on policy problems. I am inspired by her spirit of entrepreneurial problem-solving, dedication to working in the public interest, and commitment to building better institutions. Working with Jan on the Transportation Data Collaborative indelibly shaped my perspective and this dissertation.

This work would not have been possible without Mark Hallenbeck, who supported me as a mentor, domain expert, research participant, and collaborator on the Transportation Data Collaborative. Mark's centrality to transportation data in my field site made him an irreplaceable source of insight and knowledge. His warmth, humor, and enthusiasm made even the most arcane aspects of transportation data systems shine and take wing.

I was very fortunate to work with Ryan Calo early on in my program. His conviction that public policy problems in emerging technology require an interdisciplinary approach set the blueprint for my education. His influence and vision also helped to guide everything that happened after that. The University of Washington Tech Policy Lab made it possible for me to be work on deeply cross-disciplinary projects, particularly a 2015 effort in the City of Seattle that introduced me to much of the material that would become this dissertation. He taught me how to critically interrogate gaps—in words, definitions, and in the application of law and policy. His critical and incisive feedback elevated this work and supported my evolution at key points in my development as a researcher.

I would also like to thank Nic Weber. In asking Nic to serve on my committee, I knew I would rely on his rigorous standards as an ethnographer and domain expertise on the data infrastructure worlds I was studying. What I got from Nic besides these things is the gift of his support and encouragement.

I also thank Daniela Rosner, who played an essential role in my development as a field researcher. Giving me the opportunity to work with her on the Rethinking Repair project in 2014-2015 taught me about the nuts and bolts of ethnographic research; it was from her I learned how to translate a mess of fieldnotes into participant observations. Her encouragement for me to foster ties in Seattle tech activist networks also helped root me to Seattle as a field site over the years to follow.

I would like to thank Batya Friedman, who I learned so much from over the course of my PhD. Thank you to Anna Lauren Hoffmann, who inspires me deeply in the way she translates her values into action through research and teaching. Thank you Adam Moore for supporting my

exploration and helping me locate this work in the passage point between public and private as an enduring and generative set of debates.

Thank you to the activists, public employees, advocates, practitioners, and others who took the time to talk to me on the work in this study, and all those who took part in the related research projects that helped to develop it.

Thank you to Sue Glueck for the idea to fund the research that made all of this possible for me; to Emily McReynolds for taking note of my hope to start working in tech policy; and to Ginger Armbruster for all I learned from her about how data is governed at the City of Seattle. I would also like to thank Norah Abokhodair, Maggie Fesenmaier, Michael Mead Yaqub, Mariam Yaqub, Jason Portenoy, Matt Willis, and Jérôme Cattin for cheering me on.

I have saved a special space at the end here to thank my collaborators and friends, Peaks Krafft and Michael Katell, who co-founded the Critical Platform Studies Group with me as a way to pursue our shared interests. Working with them has been like being in The Beatles, where all of us are John and all of us are Paul. It's been pure joy and an honor. We have done so much. To much more!

I would like to acknowledge the inspiration I draw from Aaron Swartz, who fought to ensure that publicly funded research should be open and available to all.

The deepest thanks goes to my mom and dad, Andrea and Robert, who have encouraged me at every moment. They have supported me to follow my path as it unwound and evolved. I have drawn strength from both of you. Your integrity and kindness set a powerful example for me. Thank you for believing in me.

DEDICATION

To my grandmother, Grace Young, who, with three small children at home and a family farm to help my grandfather run, went back to the University of Iowa to finish college in 1959 and went on to earn a master's degree. She put her education to work in public service through years of teaching special education at schools and clinics. Her lifelong dedication to learning and to helping others pioneered the trail I am on.

To my mom and dad, Bob and Andrea, whose hard work and careers I am proud to say helped many people. My dad as a physician, who is loved as a warm and skilled listener, and my mom through directing healthcare research at scale as a kind leader and critical thinker.

Chapter 1. INTRODUCTION

Law scholar Lawrence Lessig (1998; 2009) famously claimed that "code is law," that is, the design of software has meaningful impacts on our lives; code can constrain our actions just as laws, norms, or markets do. Whereas Lessig was writing on how laws could apply to cyberspace, we may instead consider how code is akin to law when government relies on technologies to provide public services. For instance, the software in automated license plate reading devices police that officers use can play a role in how laws are enforced; errors or omissions in its code can shape who is subject to a traffic stop or ticket in practice.[1] However, code and data is less amenable to public oversight. Whereas laws are codified and available for all to read, in many cases, the software, code, and data behind public sector technologies are not readily accessible to public view or evaluation.

The stakes for justice are high. As scholars like Virginia Eubanks (2018) and Ruha Benjamin (2019) observe in their work, bias in software can deepen and amplify racial and socioeconomic inequalities. Such harms have been documented across many machine learning tools, such as those used for recidivism prediction (Angwin et al. 2016), face recognition (Buolamwini and Gebru 2018), and healthcare decision-making (Obermeyer et al. 2019). The potential for harm also arises in more simple datafied systems and processes, such as sampling processes used in jury selection (Randall, Woods, and Martin 2008). A burgeoning academic field focused on the fairness, accountability, and transparency of software finds that such harms result from multiple sources in the ways data is collected, code is written, models are trained, and software is used. Oversight of information systems, then, may require access to data, as well as code and models (Young et al. 2019).

This dissertation traces specific cases of data sharing between the public and private sector by tracing bids to data access as they come into conflict with intellectual property claims. At first glance, these conflicts broadly map to familiar tensions between privacy and transparency, such as a firms' interest in protecting their intellectual property and public agencies' duties to disclose public records. However, a closer look reveals a more variegated set of frictions.

[1] In a study looking at the use of automated license plate readers in Seattle, the software used to detect license plates was not capable of detecting the state a license plate was from, leading to occasional false positives in the event of an alphanumeric match (Young, Katell, and Krafft 2019).

1.1 GOVERNMENT USE OF TECHNOLOGY AND INTELLECTUAL PROPERTY

In an effort to make public services more effective, efficient, and data driven, government agencies often rely on information technologies like sensors, storage, analytics, and services they buy or contract with private companies to provide (Söderström, Paasche, and Klauser 2014; Kitchin 2014; Whittington et al. 2016). In this respect, contracts between public agencies and private firms both enable new government services and present new intermediaries to agencies' and residents' access to information about those services. This sense of "intermediation" emerged soon into my first encounter with local open data advocates and open data users (known as "civic hackers") in the City of Seattle in 2015, where I interviewed them about their hopes and concerns for Seattle's open government data program. The City had contracted with a firm called Socrata to create and manage city's open data portal.[2] As part of this agreement, an employee at Socrata became responsible for managing requests from the public for data access, such as particular file formats. As one respondent explained:

> "The concern, the fear is now the government entity can say, oh, we've opened up the data and we're done. These guys [i.e. Socrata] are now our front-end, and [the City] merely shifted it. A closed corporate model actually exacerbates [access issues], because then there's no, little ability to influence the scheduling of those projects or when data gets put up when. … I have a fear that those intermediaries will inadvertently become a larger barrier than dealing with a government agency that I can always hit with a Freedom of Information Act." (Respondent in focus group with Code for Seattle, February 12, 2015).

In this instance, the respondent had wanted access to data in a raster format, and had been discouraged by a response from the Socrata employee fielding these requests who asked, "What do you need that for?" Here, the respondent expresses that it is easier to insist on data access directly from government agencies by using freedom of information laws.

Moreover, requests by the public for visibility into government software and data can sometimes be denied or challenged on the basis of intellectual property claims by firms. Private firms regularly stake a claim over some subset of the goods and services they provide to governments; for instance, when a firm submits a bid to a public procurement process, they may designate the materials they submit as "confidential." In particular, manufacturers of software used by government may assert its code to be trade secret. Recent court cases have raised the question of how proprietary tools used by law enforcement should be evaluated and by whom—particularly

[2] http://data.seattle.gov

in cases where officials rely on an assessment made by proprietary software. For instance, in 2011 in Long Beach, California, local police used DNA testing software called TrueAllele manufactured by a company called Cybergenetics to test samples from a cold case. When they found a match in a national database of previously convicted people, the defense attorney of the man accused then filed a motion to compel discovery for access to the software's source code. Cybergenetics fought this request on the grounds it would violate its trade secret. The court ultimately prevented the code from being disclosed (Tashea 2017). Other cases demanding to audit TrueAllele's source code were denied in Ohio, Pennsylvania, and New York (Murray 2015; EPIC n.d.). Beyond genetic testing, defendants have also requested access to source code for technologies like breathalyzers, which have similarly not been granted (Wexler 2015). In all of these cases, intellectual property claims over vendors' software interfered with the ability of the public to access or evaluate the contents of their code.

1.2 Intellectual Property Claims as a Black Box

This study is inspired by scholarship on the social implications the use of proprietary technologies. The legal and sociological dimensions of this problem are explored in work such as law professor Frank Pasquale's (2015) *Black Box Society*, where the term "black box" is used to refer to technologies protected via intellectual property laws like trade secret that allow companies to prevent their data and software from being exposed under the argument that firms must protect their competitiveness. Pasquale raises critical questions about how firms dodge public accountability and government oversight through such claims. In a similar vein, danah boyd & Kate Crawford (2012) point to the extreme difficulty for researchers in gaining access to social media companies' data; they warn of the chilling effect that such limited data access has on the scope and realization of research and of the asymmetrical power relationship between firms and researchers in this regard. In this study, I focus on government agencies themselves as "users" of proprietary systems. I consider how agencies (and members of the publics they serve) are affected when firms assert some subset of their work to be intellectual property. While scholarly discourse to date focuses on *platforms* and *software*, here I focus on *data access* as one analytical cut into a larger discussion of the societal implications of "black box" software.

1.3 Conflict Between Public Records Laws and Intellectual Property

Activists have also raised concerns about the implications of intellectual property claims as they intersect with public records laws. Tension between vendors' confidentiality claims and freedom of information requests emerged early on in my foray into local tech activist discussions in Seattle. In Tacoma, Washington in 2014, a transparency activist and co-founder of the Seattle Privacy Coalition named Phil Mocek filed a public records request to the Tacoma Police Department for information about a specific surveillance device called Stingray (manufactured by a company called Harris Corporation), which was used to intercept suspects' mobile phone calls. In response, the Tacoma Police Department disclosed a document called "nda_redacted.pdf" with significant portions of the document blacked out (see Figure 2), asserting an exemption in the law for investigative records. In comparing the redactions in the documents he received to other publicly available records, Mocek found that these redactions were intended to withhold the contents of a non-disclosure agreement that the department had signed with Stingray's manufacturer, and disputed that decision in court (Miletich 2015). Mocek's 2015 appeal of the disclosed documents argued that the redactions, as well as the contents of the manufacturer's non-disclosure agreement, violated the broad pro-transparency construction of state public records law (Robinson 2015). A court agreed and granted Mocek $50,000 in damages, which he used to start a small surveillance transparency and accountability nonprofit (Ruud 2017).[3]

This finding was made possible in the context of Washington state's commitment to transparency as reflected in the scope and staying power of its freedom of information law, the Washington Public Records Act (RCW 42.56). The Act was passed by ballot initiative in 1972 and is one of the most pro-transparency state laws of its kind, requiring that Washington public agencies must disclose most records they prepare, own, use, or retain. By design, the Public Records Act is "liberally construed, and its exemptions narrowly construed" (RCW 42.56.030) so as to incline disclosure of information public employees use, prepare, own, or retain. This inclination has proved over time in court to apply broadly.[4] Over the course of my fieldwork,

[3] That same year, I learned about the case from Mocek over pizza at a monthly gathering in Seattle called Tech Activism Third Mondays. This story sparked what would become for me a set of persistent questions about the role firms play in public service provision and how governments navigate these relationships while maintaining their obligations to be transparent, fair, and accountable to the public.

[4] The work I report on here began when Jan Whittington and I interrogated the application of Washington Public Records Act to data-intensive technologies. We had worried how the Act could inadvertently expose sensitive records, given the absence of explicit exemptions for many types of record. Police body-worn cameras illustrate this

public records requests produced access to government data as a type of public records, including numerous forms of tabular data, relational databases, videos, photos, and geospatial data.

There is no privacy exemption in the Act.[5] Rather, narrowly defined exemptions to disclosure are passed by the legislature.[6] In the interest of transparency, exemptions are so narrowly defined as to require an enormous number to be individually enumerated in the statute—more than 400 by last count of the Washington Sunshine Committee, a legislative body created to uphold the transparency values of the Act. The primary effect of these provisions is that sensitive information cannot be withheld from disclosure, even when it provides personal information about identifiable individuals.

Firms' efforts to prevent data disclosure were shaped by the provisions of the Act. In a process known as "third party injunction" (RCW 42.56.540 and RCW 42.56.520(2)) agencies are empowered to notify third party firms when the agencies have received public records requests implicating information about them, thus giving them the opportunity to file an injunction in court. The ultimate disposition of the contested records depends on a court determination about whether disclosing the records would violate firms' trade secrets. In many cases, firms filed third party injunction as a means to protect their data from release, citing trade secret. In this sense, the Act itself indelibly shaped the case studies I report on by making intellectual property claims a salient strategy to attempt to prevent data from being released.

In the case studies to follow, I find that an underappreciated corollary of public access to government data is that the Act also *mediated* government agencies' own access to data from its vendors, partners, and third party firms. That is, public agencies I researched in Washington encountered challenges in gaining access to firms' data due to firms' concerns over exposure. This problem reflects the law's intention for records available to agencies and to the public to be largely

collision of interests in information access and control. When the Seattle Police Department adopted body worn cameras, they were intended to be a tool for police transparency and auditability In Seattle in 2015, a local activist, known for months only as "the anonymous requestor" had created a bot (i.e. software script) to automate submission of public records requests for all of the body camera video from police departments in multiple jurisdictions across the state – since there was no explicit exemption for body camera footage in the Act, the Seattle and Bellingham police departments disclosed unredacted videos to the requestor-- who posted them on YouTube, with dire privacy implications.

[5] The Act defines privacy narrowly as information whose disclosure would be both "highly offensive to a reasonable person" and would "not [be] of legitimate concern to the public"– in practice, courts find few public records requests that rise to this standard (RCW 42.56.050).

[6] Exemptions are also interpreted strictly. For instance, an exemption for public employees' residential addresses does not apply to closely related information, such as geolocation datapoints of public employees' residences.

coextensive by design. While efforts to reform the Act in a way that will protect individual privacy and facilitate agencies' access to data have begun to formalize during the time of this writing in 2020, public agencies' frictions in gaining access to firm data appeared throughout the fieldwork I conducted for this dissertation from 2015-2019.

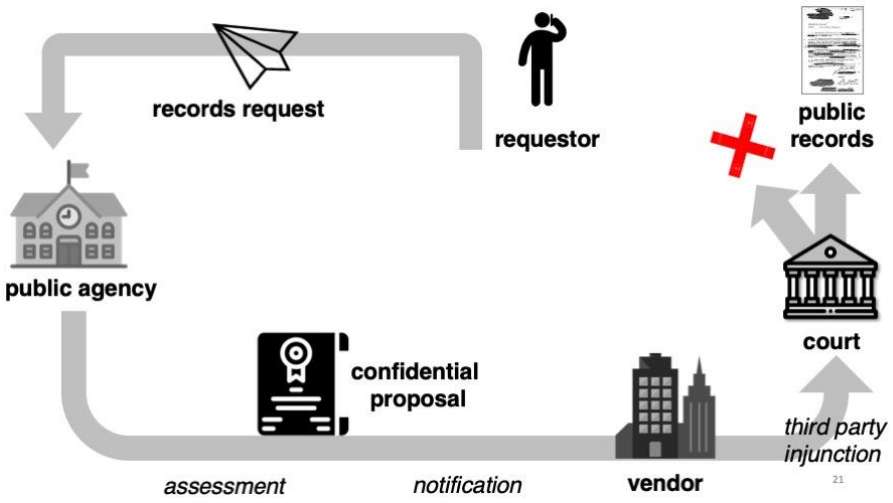

Figure 1. This diagram illustrates third party injunction, a process in which public agencies are able to provide vendors with notification when a records request implicates their own interests. For instance, if an agency receives a public records request for proposals marked "confidential" from bidding firms, it may use third party injunction as a process to provide notice to bidding firms that the records are to be released. The vendor files an injunction in court, which makes a determination about whether records are released as public records or withheld.

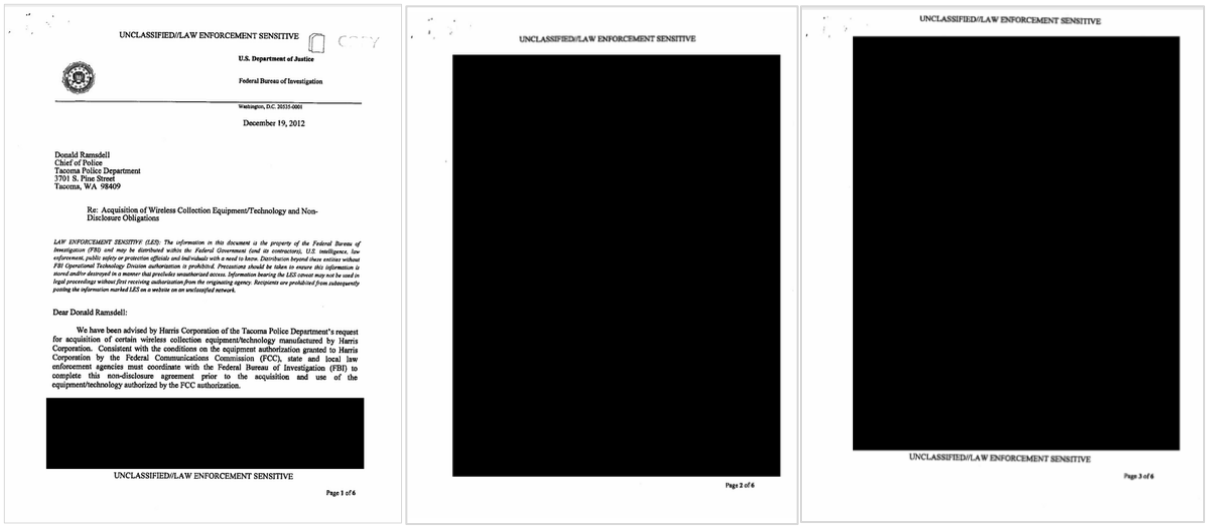

Figure 2. A document disclosed by the Tacoma Police Department to activist Phil Mocek, "nda_redacted.pdf"

1.4 Cross-Sector Data Sharing in the Transportation Sector

I locate my analysis in data collected in the transportation sector, which is increasingly the locus of innovation, outsourcing, and public-private partnerships.[7] In the United States, policy directives that encouraged "data driven" transportation policy date to at least the early 1980s, when the Federal Highway Administration was interested in statistical methods for monitoring traffic.[8]

[7] There are numerous applications for data collected in transportation. In the 20th century, transportation data was often collected by hand; for example, government employees would collect information about traffic volumes by employing people to count cars at intersections to identify peak traffic times. As the cost of data storage and sensor hardware decreases, it becomes more cost effective for to collect and store granular, pervasive, and continuous information. For instance, transportation planners are interested in data that could be used to drive more efficient stoplight timing, such as real time information about traffic volume, counting how many people board a bus or train, or automatically deducting tolls. Data also permits one application to be interoperable with another; for example, data to collect parking payments can also enable an application that provides real time information about parking availability. In addition to the primary uses of data, it has secondary uses to evaluate programs and plan future services.

[8] Most transit agency data collection in the 1980s-1990s time period was for making decisions about resource allocation and prioritizing maintenance. A major driver encouraging multiple transit agencies to adopt this shift occurred in 1998, when the US DOT initiated the Intermodal Surface Transportation Efficiency Act of 1991 created reporting requirements for departments of transportation to report indicators on the state of highway pavement, bridges, highway safety, traffic congestion, public transportation facilities, and intermodal transportation facilities, and systems. The law was met with pushback from local departments of transportation because it required data management systems that transportation agencies did not have; the agencies "fought them tooth and nail because they didn't want to put money into put money into [buying] systems" (interview, transportation researcher Feb 18, 2018).

Today, data is central to many aspects of transportation planning such that as a sector, effective cross-sector data sharing and collaboration plays an increasingly important role in present and future visions for the field. Public transit agencies would like data sharing and access to enable new programs, enforce local laws, and aid in transportation planning-- but in practice key efforts have failed due to data sharing frictions.

Private firms in the transportation sector hold a growing share of data that agencies could ostensibly use to help guide planning, including widely adopted services like car share (Jorge and Correia 2013), ride share (Ma, Zheng, and Wolfson 2014), bike share (Zhang et al. 2016), prediction apps for public transportation (Ferris, Watkins, and Borning 2010), and routing apps (Brock et al. 2018). Notably, in the absence of access to firm information, transportation researchers spend considerable resources simulating it, as evidenced by work modeling the supply and distribution of rideshare vehicles (Jorge and Correia 2013). Under pressure from the public to make data available, firms like Uber and Waze have provided some version of their data via feeds and dashboards. Nevertheless, these efforts have failed to realize the scope or potential of data sharing for transit. At the time of this writing in 2020, the aftershocks of a high-profile court case between Uber and the Los Angeles Department of Transportation attested to enduring tension in this regard:

> *"In an LA City Hall hearing in January, a lawyer representing Uber declared: 'It is surely no exaggeration to say that what LADOT is proposing here resembles the world of a dystopian novel, like 1984 or Brave New World, where the government tracks its citizens in real-time about where people were, where they are, and where they are going.' While the Orwellian comparison has so far failed to convince the courts, the case has raised important questions about how cities use data and what limits companies are prepared to go to [in order] to stop them" (Carey 2020).*

Here, this quote equates government access to data with privacy and surveillance concerns, illustrating the powerful set of competing interests and uneasiness in conflicts over seemingly prosaic questions of data access and sharing.

1.5 Contested Claims to Data Sharing and Access

In this dissertation, I examine cases of claims to data access and sharing in which public claims to data and code were contested by firms citing trade secret. This focus on contestation, breakdown, and failure takes a cue from Susan Leigh Star's (1999) methodological tools for conducting ethnographic research on infrastructure: "The normally invisible quality of working infrastructure becomes visible when it breaks: the server is down, the bridge washes out, there is a power blackout" (p. 382). In this regard, brokenness and conflict in both cases illuminate the multiple factors at play, helping to move beyond dichotomistic tensions such as that between public records and trade secret to explore the claims that defy easy categorization. For instance, both public agencies and firms asserted "data ownership" over the same information in Chapter 5, whereas in Chapter 4, even discrete and uncontested data ownership claims were not determinative of the agency's access to data—rather, technological lock-in and practical frictions played a more important role.

Even as my analysis focuses on the shape of specific data sharing requests and subsequent developments, data requests themselves are often tied to larger, public interest-driven accountability and oversight purposes; it is these goals that are at stake when access to firm data is denied. Each case study reports on the concerns that motivated and underlie data sharing and access requests. More specifically, in data requested of a vendor in Chapter 4, agencies (and the employers who bought fare cards from them) wanted longitudinal data for use in transportation planning. In the same case, an activist requesting software wanted to establish whether the public is allowed to inspect software used in the provision of public services. In Chapter 5, public agencies vied for data access in order to assess the promise of new rideshare models for expanding transit, evaluate the equity of service provision, and audit how public funds were being spent. The same case also reports on how the City of Seattle had made data sharing mandatory as a condition of its permit to operate; data was intended to enable the city to enforce permit requirements and better account for the impact of rideshare usage on local infrastructure. Arguably, frictions to data sharing and access impeded these public interest objectives.

1.6 RESEARCH OVERVIEW

In this work, I trace claims to data sharing and access in cross-sector information systems contracts. Through ethnographic case studies, I report on data sharing as a site of contestation over data access and control mediated through the technical, legal, and policy factors. I devote particular attention to different agencies' claims and the rationales that support them, inspired by Gal (2002) who reads the terms "public" and "private" as always inherently normative. In this work, I examine claims about data as public or private property more closely. Theoretically, I locate my analysis of these contested data flows within a broader scholarship of ambiguity and political struggle in the passage point between public and private, as my cases trace data moving between across sectors-- never fully public nor fully private. In its high-transparency construction, the Washington Public Records Act emerged as a key factor in both of my case studies, particularly as it came into conflict with firms' interest in withholding data on the basis of trade secret.

I used ethnographic methods in this work. The promise of ethnographic work is to be able to provide high fidelity accounts of phenomena that seek to render different stakeholders' perspectives. I draw on an actor network theory in tracing data sharing relationships outward, beginning with the initial contract between each public agency and private firm, then following data sharing and access requests as they connect the cases to other stakeholders. The resulting two cases document the set of technical, legal, and policy factors mediating these requests. The first case study reports primarily on the contracting relationship between a regional transit partnership and its fare card vendor. It also describes the work of a local activist who sought access to the vendor's proprietary software. The chapter illustrates that public agency ownership over data did not ensure the agency's free or frictionless access to it. The second case study reports on a would be partnership between a transit agency and a rideshare company to pilot subsidized trips to transit hubs. The chapter illustrates that the data sharing needed to facilitate this program was considered risky to the firm given the visibility it would give public agencies, its competitors, and the public into its operations; the partnership failed in large part based on the firm's reticence to share data.

I focus on data ownership because it emerged as a common rationale behind data access claims and denials.[9] Examining how "data ownership" is enacted on the ground in both case

[9] Data ownership as a common refrain was shaped in part by the construction of the Washington State Public Records Act, in which the exemption used for intellectual property is a reliable means to protect data from disclosure) (see Appendix E).

studies, I find that it is intended to exert control over who is allowed to access data, and in this regard is associated with good data governance and rights protection. However, these claims are thrown into question when multiple parties assert ownership to the same data; the promise of ownership as a means to control data is not borne out in practice. Indeed, my findings indicate that data ownership is not dispositive of outcomes with respect to who can access or control it. Instead, mediating factors like technological frictions prevented owners from accessing their own data; owners were at the same time "locked-in" to their existing system provider, impeding their ability to make autonomous decisions about where to put data they owned. Finally, data ownership was not necessary or sufficient in order to claim data access or control.[10] I close by bringing these lessons learned as they relate to the use of data ownership as a policy proposal, an idea which has recently gained ground in digital rights policymaking and discussion.

Research questions:

My first aim is to describe the policy goals that guided agencies in their data access needs, and the technologies or firms that were enlisted to collect this data. My second aim is to focus on areas of data access friction or breakdowns.

- *RQ1*. What information systems and services are private firms contracting with governments asked to provide?
- *RQ2*. What roles do data sharing and access play in public agency relationship with private sector contractors?
- *RQ3.* How do technical, legal, and policy factors mediate data access and sharing in public sector contracting for information systems and services?
    - What factors play a role in affording or inhibiting public agencies' access to data about their services?
- *RQ4.* How do different actors assert or deny data access and sharing requests? Using what rationales?
    - How effectively do these rationales allow actors to assert or deny data sharing and access claims?

[10] The observation that multiple parties can assert reasonable claims to the same data motivated the creation of the University of Washington privacy-preserving data sharing intermediary, the Transportation Data Collaborative, where as a Research Assistant I had privileged access to stakeholders and their data sharing problems.

- *RQ5.* How does data ownership function as a rationale to assert or deny data access and sharing requests? How effectively does it do so?

My research questions set out to describe cross sector contracting relationships; the data shared or withheld between the parties involved; the larger legal, policy, and technical constraints for this sharing; and the rationales each party uses to use to assert or disclaim access to data.

Among the dominant rationales used in asserting claims to data access and control, I focus on data ownership as a key theme emergent in my analysis. In part, the prevalence of data ownership in my findings is tied to the concession for intellectual property claims in the between the Public Records Act, under which third-party notice (to firms asserting data as confidential or trade secret) is the primary means by which firm data is withheld from public release. This theme also emerged many times over the course of my fieldwork, be it in meetings with former Governor Gregoire with the Transportation Data Collaborative (who asked us, "Who owns the data? Does the University own the data?") in the office of her organization Challenge Seattle, to the assertion of data as university intellectual property as part of the legal strategy we ultimately chose for the Transportation Data Collaborative.

Chapter 1: Introduction

This chapter motivates my study as exploring contested claims to data access and control in government use of information systems. Government adoption of information systems that collect identifiable data often engenders familiar value tensions between privacy and accountability. When governments contract with third-party firms to provide these systems, firms may also mediate different actors' claims to data when they assert some part of the system or data to be proprietary. Using participant observation, interviews, and document analysis, this work explores these tensions in two situated case studies of contested data sharing in two contracting relationships: (i) a vendor, and (ii) a public-private partnership; I ask, "What technical, legal, and policy factors mediate data access and sharing in public sector contracting for information systems and services?" and "How do different actors assert or deny data access and sharing requests? Using what rationales?" These cases are located in the transportation sector, the site of increasing cross-sector innovation. I find that even as agencies assert that they own the data that results from contracting relationships with vendors and partners, in one case, data access claims were encumbered by the practical constraints and costs of data export; in the other, data sharing was

stymied by firms advancing restrictive data ownership claims. Together, my work provides empirical reflections on larger discussions weighing data ownership as a strategy for data protection and argues that it does not provide a solid foundation for meaningful data control.

Chapter 2: Literature Review

The case studies I report in this dissertation take place at the passage points between the public sector and the private sector, and in the contested status of documents as public information or private information. Here I examine work that theorizes the public/private dichotomy to explore ambiguity at the intersection between the two. The work I review in this chapter problematizes neat distinctions between public and private as they are applied to matters in practice. Rather, this work theorizes the multiple types of entities that exist somewhere between the two. As it applies to information, the public/private distinction is similarly fraught; the status or character of information as public or private shifts with respect to context and is a site of active contestation. As my work focuses primarily on the conflict between public records and trade secret, I review theory on what constitutes public information, private information, and related concepts like intellectual property and data ownership. In closing, I describe how negotiating the status of information in interstitial spaces in government data can inform larger debates on accountability in proprietary systems. The core finding of this literature review is that the status of institutions and information at the boundary-lines of public and private are fraught and defy easy categorization; this theoretical finding supports my grounded inquiry into how different actors contest claims to data in cross-sector data sharing in practice.

Chapter 3: Background and Methods

In this chapter I provide background as to how I came to work with local public agencies on problems related to data sharing and governance and the salience of the Washington Public Records Act for local transparency, privacy, and data sharing concerns. I describe how my participant observation on a university-based data sharing intermediary effort with Prof. Jan Whittington exposed me to the broader context for this study as a collective action problem, and in turn informed my case selection for this work. While the confidential nature of the discussions behind the Transportation Data Collaborative necessarily limited the information I could include from that work in this study, the centrality of my work on that project to this dissertation gave me access to the confidential perspectives of key stakeholders and in turn shaped the questions posed

in this study. I go on to describe my methods as ethnographic, taking cues from the methodological tactic of documenting breakdown to illuminate the constitutive elements of an information system following Star's ethnography of infrastructure approach, and van Dijck's focus on the political economy of information services in my attention to law, policy, and contractual agreements.

Chapter 4: Case #1 on the Vendor Relationship between ORCA Agencies and Vix

This chapter is a case study reporting on fare card transaction data collected by a vendor called Vix Technology ("Vix") on behalf of a partnership of seven transit agencies called ORCA. The purpose of this chapter is to explore the frictions that ORCA partnering agencies encountered in obtaining access to data that putatively belonged to them. While the initial agreement between the partnership of agencies and the vendor specified that the vendor owned the executable code underlying the fare card system, the agencies contractually *owned* fare card data. Nevertheless, constraints imposed by the original system design ultimately required the agency to pay additional money in order to secure access to this data. I describe how these frictions contributed to the agencies' decision to replace their existing "single provider, turn-key" contract with a more interoperable, modular "Next Generation" vendor ecosystem. I go on to explore the other actors who wanted access to ORCA data (e.g. institutional customers, law enforcement, transportation planners) and the ways their access to data is negotiated through the Public Records Act and ORCA agency policy. I also describe how one records request for Vix software was denied on the assertion it would violate the vendor's trade secret. Taken together, this case study illustrates the multiple ways that ownership and access claims are asserted over data and code (by agencies, by firms, by a vendor, by activists) and how these claims are negotiated under constraints imposed in practice by technological frictions, law, and policy.

Chapter 5: Case #2 on the Would-be Public-Private Partnership between King County Metro and Lyft

In this chapter, I investigate how agencies access data in a public-private partnership model. Specifically, I focus on a pilot partnership at King County Metro to subsidize a rideshare company to expand service to public transit hubs, one result of a federal initiative to promote public-private collaboration called the Mobility on Demand Sandbox. I focus on the conflict that arose in negotiations with initial project partner Lyft around data sharing, and how these disagreements led the agency to replace Lyft with another company, Via, before the project could launch. From one

perspective, the Washington Public Records Act played a key role in constraining agency access to rideshare company data, which Lyft claimed had contributed to the delay and failure of the initial round of contracting. From another, rideshare companies' reluctance to share data goes to deeper questions about an unwillingness to expose their business to greater visibility and the possibility of further regulation. Here, data sharing conflicts illuminate frictions that persist even during a turn toward greater cross-sector collaboration advanced under the banner of "mobility on demand" partnerships. I explore how the term "mobility" collapses distinctions between publicly and privately-operated transportation, and how programs to promote it (in their various forms) necessarily require greater cross-sector data sharing.

Chapter 6: Conclusion

I reflect on contested claims for data sharing and access in government contracts, with a focus data ownership as a common rationale advanced my multiple parties when negotiating data control. I set out here to characterize how data ownership is enacted in practice, that is, made meaningful in its situated context. A grounded perspective on data ownership (and how it relates to access and control rights) can be instructive in this moment when many advocates and policy proposals are calling for ownership-based regimes to data protection, remuneration, and control/"data dignity." Here I compare the impacts (or lack thereof) that the notion of "ownership" effects in the context of each case study and reflect on the implications of these findings for broader policy debates. I find that data ownership is intended on the ground to provide control over who is allowed to access data, and in this regard is understood by many to signify good data governance and rights protection. However, in my case studies I found that the promise of ownership as a means to control data is not borne out in practice. Ownership claims are contested when multiple parties assert it over the same data. Indeed, my findings provide examples in which mediating factors like technological frictions prevented owners from accessing their own data or were "locked-in" to an existing information system that impeded their ability to make autonomous decisions about where to put data they owned. Finally, I argue that my findings call into question the promise of contemporary policy and advocacy efforts to give individuals ownership over data about them, such as in recent state and federal bills (some of which have been adopted into law) or part of prominent political platforms.

1.7 CONCLUSION

In this chapter, I have introduced my study as an exploration of data sharing and access claims in government contracting for information systems. These relationships are the site of multiple forms of conflict—including familiar tensions between intellectual property versus transparency, privacy versus public benefit, and competitive advantage versus accountability. At the same time, a close and situated look at claims made by actors in context produce novel understandings of how these terms carry meanings in practice; I particularly focus on data ownership as a common rationale behind both assertions and denials of data access by public agencies and private firms. In both of the case studies to follow, the Washington Public Records Act played a key role in mediating data access. I have also provided a justification for focusing this work in the transportation sector, where an increasing focus on "mobility" service provision has accelerated public agency interest in cross-sector collaboration, where data plays a central role in the deployment, evaluation, and future vision of these efforts. In the next chapter, I will explore how the boundary between the "public sector" and the "private sector" becomes enmeshed in hybrid institutional configurations, and how data too is ambiguous with respect to its publicity, privacy, and status as property.

Chapter 2. LITERATURE REVIEW

The work I review in this chapter problematizes neat distinctions between public and private as they are applied to matters in practice. I begin with this distinction because the case studies I report in this take place at the passage points between the public sector and the private sector, and in the contested status of documents as public information or private information. In this chapter, I examine work that theorizes the public/private dichotomy and which explores hybridity and ambiguity at the intersection between the two. I follow notions of public and private from their etymological and conceptual roots through to their dichotomization and application. I focus particularly on the "public versus private sector" and "public versus private information." The distinction between public and private sector is at times ambiguous given the multiple types of entities (government vendors, public-private partnerships) that cannot fit neatly between the two. The boundaries of the state are continuously re-drawn in practice over time and through the associations traced by government in its various contracting relationships. Similarly, the status or character of information as public or private shifts with respect to context and normative claims. I review theory on what constitutes public information, private information, and related concepts like intellectual property and data ownership to show how even intellectual property laws acknowledge a public interest in some access to proprietary information.

These theories are the conceptual scaffolding on which my case studies will trace data in transit across the public-private sector boundary. The core finding of this literature review is that the status of institutions and information at the boundary-lines of public and private are fraught and defy easy categorization. Key work by figures like Susan Gal (2002) understands the terms "public" and "private" to not inherently describe things in the world, but instead to be used to make

normative claims. That is, specific actors use them to assert what should be considered public or private. This theoretical departure point supports my inquiry into the normative claims different actors make to claim and withhold access to data. It invites a focus on situated empirical cases as these claims are asserted in practice by actors given their various interests and value commitments.

2.1 THE PUBLIC-PRIVATE DISTINCTION

The public-private distinction figures prominently in social theory and practice. However, some scholars find that the distinction breaks down on empirical examination-- even as they affirm its enduring value as a conceptual and ideological tool. This chapter provides a cross-disciplinary survey of influential works defining the boundary between the public and private sector, that between public and private information, and ambiguity that arises at these intersections. Although a systematic review is beyond the scope of this chapter, I include a number of influential works from geography, sociology, political theory, philosophy, law, and anthropology that take up the public-private distinction as a starting point for my discussion of public-private data sharing.

2.1.1 *Defining notions of the public*

The many connotations of the word 'public' primarily cohere around three interlinked ideas; access, communality, and representation. First, the word connotes visibility, openness, accessible to others for inspection, or generally known by them (Pitkin 1981; Okin 1989). This valence also extends to a public ethos, which is associated with democratic values like openness, transparency, and fairness (Newman and Clarke 2009). For example, places are public when anyone may enter (Lewis and Wilson 1877).

The public also signifies what is shared among people as a whole. Etymologically, the word comes from Anglo-Norman meaning 'of the people' as a single communal body, or someone authorized by, serving, or representing them (Oxford English Dictionary n.d.). Something public "has no immediate relation to any specified person or persons, but may directly concern any... members of the community without distinction" (Lewis and Wilson 1877, 22–24). Discursively, this notion of the public ties it to the people as citizens of a nation or other shared identity and forms the basis for collective provision of goods and services (Newman and Clarke 2009).

Hirschman (1982) similarly describes it as any action taken by a citizen in civic or community affairs. In its official capacity as representing the people, something 'public' is also impersonal, distant, or formal (Pitkin 1981). Any actions taken by a magistrate or government official are public (Lewis and Wilson 1877). In this discursive formation, the public comes to be equated with the public sector and the state.

2.1.2 *Defining notions of the private*

Etymologically, to be private is to withdraw from the larger group; its early uses connoted separation, restriction, and seclusion. It shares a root with the word privilege as well as deprivation, which both share a sense of being apart, closed to outside access, or hidden (Pitkin 1981). Most texts trace our modern conception of the "private" to the Ancient Greek concept of *oika*, the household, the basic unit of production and the root of our word economics. Just as the economic was considered to be non-political in Ancient Greece, the home was also outside the political realm (Arendt 2013). As it relates to the household or home, the private is associated with bodily, material concerns, such as intimate relationships, reproduction, and physical labor (Arendt 2013). Thus, the private has two core elements; one that is solitary or related to the family, and another concerned with material welfare and the market (Hirschman 1982; Gavison 1992). This discursive formation underlies much of the liberal tradition.[11]

2.2 THE PUBLIC-PRIVATE DICHOTOMY

The conceptual roots of the private/public dichotomy are also traceable to ancient Greece; virtuous men had to choose between *vita contemplative*, a private life associated with solitary

[11] The public-private dichotomy corresponds to a tenet in liberal philosophy that there is a limit on what the government should do to an individual, given his or her natural rights (Blomley 2005). Classical liberals were concerned about public institutions subjugating free, rational, equal individuals (Mill 1869; Fairfield 2005), and sought to provide the private sphere a measure of freedom from government intrusion (Gavison 1992; Mnookin 1981). "The liberal ideal of the private holds that, so long as the public does not interfere, autonomous individuals interact freely and equally. Privacy is the ultimate value of the negative state" (MacKinnon 1989, 190). Put another way, the liberal tradition constrains public policy to protect private rights (Benn and Gaus 1983). In this way, a contemporary understanding of the public-private dichotomy has been deeply shaped by liberal philosophy. Feminist scholars have observed the extent to which this freedom has permitted the household to be controlled by domination and force (Arendt 2013). Indeed, the distinction has been explicitly gendered since its inception and conceptual antecedents in Ancient Greece; only free, male citizens could participate in the *polis*, and women were relegated to domestic life (Fairfield 2005). Similarly, moral philosophy in Ancient Greece esteemed the virtues of public, political life and participation. As public life and persons were valorized, private persons were considered as lesser or irrational (Elshtain 1981; Gal 2002).

contemplation, and *vita activa*, involvement in public service (Hirschman 1982; Arendt 2013). Gal (2002) argues that the dichotomizing impulse of the nineteenth century further elaborated this binary to strands of discursive alignment between public vs. private, the community vs. individual, rationality vs. sentiment, money vs. love, and solidarity vs. self-interest. Klein (1995) highlights the influence of the structuralist tradition in considering conceptual binaries to be foundational to social worlds.

Following this dichotomizing impulse, essentialist arguments have also asserted fixed, universal experiences of publicity and privacy. To date there is considerable disagreement as to a definition of either that adheres across space and time.[12] Instead of deciding on an intrinsic quality to each term, Solove (2008) and Klein (1995) assert that the public/private dichotomy follows Wittgenstein's (1953) observation that some concepts cohere around a set of "family resemblances." However in themselves, each term does not signify unambiguously. Weintraub (1997) finds that those who use terms like "public" and "private" are not usually clear about which of the wide array of aforementioned meanings they intend, and "without quite realizing it, mean several things at once" (p. 2).

Importantly, the terms "public" and "private" can be intended in both descriptive and normative ways. Gavison (1992) cautions that because they are used both descriptively or normatively, the description of something as "private" does not necessarily have normative force. In contrast, Gal (2002) reads both terms to be always inherently normative; the terms "do not describe the world in any direct way, they are rather tools for arguments about and in that world" (p. 79). Weintraub (1997) also reads them as a product of ideological commitments, specific sociohistorical context, and set of assumptions. These latter views demand relational, dynamic accounts of public and private that do not consider them to be fixed properties of things, but rather assertions or claims made from particular situated perspectives.

[12] Previous work seeks to recover the understanding that the history and construction of the public-private distinction is steeped in Western norms and culture. Lisa Drummond (2000) draws on nine years of ethnographic research in Vietnam to trace the public-private boundary in contemporary Vietnamese urban life. For example, public space hosts many activities that are conventionally considered to be private, such as families cooking and eating, entrepreneurs selling wares, or couples making out. The state also takes an active role in domestic matters, such as family planning and women's role in the home. Rather than dismiss the public-private dichotomy in the face of these empirical departures from its conventional form, she argues that it "still retain[s] substantial descriptive power at an everyday level" once cultural and historical context is taken into account (p. 2377).

Metaphorically, the public-private dichotomy is often described in spatial and geographic terms, especially as a boundary (Sennett 2017), but also as a line or border (Stark 2010), and a divide (Blomley 2005). Other approaches de-emphasize a strict contrast in favor of more nuanced distinctions; for example, that the terms are relative (Pitkin 1981); continuous (Benn and Gaus 1983); co-constructed (Bailey 2000); fluid (Blomley 2005; Marx 2001); fractal, that is, nested within one another and shifting with respect to the perspective of the viewer (Gal 2002); and distinct but interrelated (Gavison 1992). More recent scholarship on the public-private dichotomy has opted to highlight its dynamic and relational nature.

2.2.1 *How to draw the line*

The public-private dichotomy operates in a wide variety of contexts; it is difficult to demarcate a boundary that extends across them all-- compounded by the challenge that the terms can be used both descriptively and normatively. Instead of arriving on a definitive characterization, some have attempted to systematize the multiple dimensions of the distinction. These dimensions tend to converge on common themes. Benn and Gaus (1983) map the dichotomy in terms of access (to something physical, activities, information, or resources), agency (on whose behalf someone is acting), and interests (whether gains and losses are individual or communal). Pitkin (1981) highlights similar dimensions: access, impact, and governance. Gavison (1992) structures the concept as related to accessibility vs. inaccessibility, interference vs. freedom, and society vs. individuals; she also provides for what she calls "complex" meanings--the uses of the term that do not fit within such rigid structures. Under this construction, violating privacy in one sense does not necessarily violate it in another.

	Access	**Agency**	**Interests**
Public	Accessible	On behalf of all	Gains/losses are communal
Private	Inaccessible	On behalf of few	Gains/losses are individual

Table 1 Three dimensions of public vs. private, Benn and Gaus 1983

	Access	**Intervention**	**Interests**	
Public	Accessible	Interference	Society (groups)	Complex meanings
Private	Inaccessible	Freedom from interference	Individuals	

Table 2 Three dimensions of public vs. private, Gavison 1992

	Access	**Impact**	**Governance**
Public	Wide, general access	Broad effects for many people	Collectively managed
Private	Limited access, seclusion	Affects few people	Control by a few

Table 3 Three dimensions of public vs. private, Pitkin 1981

2.2.2 *Objections to the dichotomy*

Many scholars have criticized the public/private dichotomy for its failure to accurately characterize the social world. A predominant critique of the public-private distinction is that it is a distinction without a difference; it obfuscates similarities between both spheres (c.f. Wedel et al. 2005). This criticism obtains in empirical scholarship from many domains, including scholarship troubling the discursive formation of the public versus private sector, and public versus private realms (Klein 1995; Foucault 1978). Ellickson (1982) draws on observations of private entities--homeowners' associations--to liken them to cities; for example, their membership can be coerced with regulation in the form of rules, and taxation in the form of monthly dues. Stark (2010) provides more recent evidence from private, gated communities that argue that their quasi-public functions in providing security, roads, and other amenities to residents should allow them to deduct their homeowners' association dues from their state and local taxes. Mnookin (1981) says critiques like these, drawn from the dissolution of the distinction in practice, to "obliterate" its usefulness (p. 1435).

Feminist scholars have critiqued the public-private dichotomy from within and without--that is, by disputing the content of the categories 'public' and 'private,' as well as the distinction altogether. Drawing attention to the dichotomy's supposed inevitability, feminist work instead situates it in its historical context as a product of the liberal tradition (Fraser 1998). While state

restraint from intervening in the private sphere has been intended to protect bodily integrity and personal freedom, feminist theorists like Mackinnon (1989) and Romany (1993) indict its tacit allowance of violence against women in the home. In response, feminists cohered around the rallying cry that the 'personal is political,' that is, that the most intimate affairs of one's life are product of power relations-- as well as the battleground they are fought upon (MacKinnon 1989), whereas the liberal notion of privacy assumes that individuals should be given autonomy to act freely and fairly, this view elides realities of domination and abuse. What falls on each side of the boundary thus becomes an urgent political question, given that "not everyone stands in the same relation to privacy and publicity; some have more power than others to draw and defend the line" (Fraser 1998).

Other scholars dismiss the dichotomy on the grounds that it has become conceptually muddled and ineffectual. Legal philosopher Duncan Kennedy (1981) asserts that the public-private distinction posits the distinction's "internal breakdown, incoherence, and ritualized maintenance" (Turkel 1988, 819), such that it can no longer be intuitively and consistently applied. Kennedy (1981) traces the failure of the dichotomy into six phases; the first being its application to edge cases with strong disagreement on both sides; followed by the development of intermediate terms, in which a given matter is neither fully public nor fully private; then a general collapse of our ability to meaningfully differentiate one category from another without internally contradicting ourselves; and finally the effort to resuscitate the distinction by mapping it to a continuum. Kennedy argues that this process leads to the politicization of definitional arguments; and ultimately, the resemblance of extreme ends of the spectrum to each other. This latter phenomenon (which he calls “loopification”) re-asserts the idea that the distinction fails in practice—but instead of failing in the interstitial space between the two concepts, notes its failure at either extreme of the spectrum.

2.3 Alternatives to the Dichotomy

Moving beyond what they consider to be a flawed paradigm, some scholars have provided alternative schemas to describe the social world.

2.3.1 *Trichotomy: public/private/society*

In response to this ambiguity, other scholars introduce a trichotomy between public, private, and civil society (Weintraub and Kumar 1997) or "the social" (Arendt 2013). The social includes commercial and economic activities that are increasingly in the public realm, such as corporations, clubs, churches, and professional organizations; Arendt intended it to describe a "curiously hybrid realm where private interests assume public significance" (Pitkin 1981). Benn and Gaus (1983) enjoin that while this trichotomy may be illuminating, it has not been adopted by the wider culture, which instead seeks to redefine the dichotomous categories of public and private.

2.3.2 *Entanglement*

Others have argued that both dichotomizing and trichotomizing accounts overlook the extent to which these realms are entangled in practice. Disparate interests constantly struggle to align themselves along "a plurality of potentially conflicting logics...[which] produce strange confluences, alliances, antagonisms, and paradoxes" (Newman and Clarke 2009, 17). A tradition in sociology called “culturalism” argues that the state is "necessarily intertwined with non-state actors" (Mayrl and Quinn 2016; Mitchell 1991). This entanglement becomes more visible in the specificity of particular cases; for example, Freeman (2000) argues that the public and private sectors are mutually constitutive and interdependent by examining the vital role that the private sector plays in healthcare, prisons, and the formation of regulatory standards under administrative law. In the same vein, scholars in sociology tracing the associations between government and its contractors or private partners do not presume particular boundaries around the state, but trace state boundary formation as a dynamic process, which they call “associational policymaking” (Mayrl and Quinn 2016) (see Section 2.5.4).

2.4 Continuing Salience of the Dichotomy

While empirical research has undermined the validity of the public-private dichotomy in practice, scholars note the persistence of this conceptual distinction as a tool for making arguments about the world (Gal 2002). Gavison's (1992) corrective to these critiques is to ascribe the terms to phenomena in a more deliberate way, rather than dismiss the dichotomy entirely. In contrast,

Kennedy (1981) cautions of that the public-private distinction can no longer hold weight as a serious "description...explanation...or justification of anything" (p. 1357). Years later, Beswick (2016) responds to Kennedy through parody in writing “the decline of the fish/mammal distinction” to echo Kennedy’s critique of the public/private distinction, noting for example, "there are fish-like creatures and there are creatures that bear less resemblance to what we would traditionally call a fish"—ultimately urging that we ought not "abandon use of the term altogether" (p. 94-95). Beswick's argument is that such categories help us to make more sense of the world rather than less.

Previous work has also defended the dichotomy on the grounds of the different rights and expectations that flow therefrom. From an organizational studies perspective, Wamsley and Zald (1973) find argue that while some believe public and private organizations are similar in all meaningful ways, the public sector is indeed subject to an idiosyncratic set of constraints. In particular, governmental authority ultimately rests on coercion through its monopoly on the use of force. It also symbolically represents society as a whole. "From these fundamental features flow definitions of membership, rights, expectations and obligations to the state and its agencies. Citizens...feel they have different 'rights' and 'expectations' with regard to the FBI than they have with General Motors” (p. 63). This elevated set of increased rights and expectations raises the stakes for what matters are considered to be public.

2.4.1 *Consequences of the boundary*

The definitional boundary between public and private matters for applying law and policy. Among other things, for a matter to be “private” is to presume it is largely outside the "legitimate bounds of government coercion and regulation," save for the purpose of enforcing contracts and preventing fraud (Mnookin 1981, 1430). The content of what is considered public or private thus becomes a political battleground for the scope of legitimate state action.[13] In the United States, issues like

[13] Some scholars have remarked that rather than the line between public and private guiding how the law is applied, it is drawn to justify the exercise of asymmetrical power relations. Klare (1982) examines the role of the public-private distinction in workplace disputes to argue that it used to dominate and atomize less powerful people in order to prevent collective or state action. He unpacks specific doctrinal examples in labor law to highlight ways in which the same entity, like the workplace environment, is variously characterized as public or private. Ultimately, he argues that it would be a grave misconception "to imagine that legal discourse or liberal political theory contains a core conception of the distinction capable of being... applied in a determinate manner to concrete cases" (p. 1361).

birth control, abortion, environmental regulation, and eminent domain have been contentiously debated as being matters of public rule or private individuals. Political parties and advocacy groups have organized along these fault lines, deploying resources to defend a particular vision of how the line should be drawn (Mnookin 1981). Critical legal geographer David Delaney (2010) focuses on how these legal distinctions come to be spatialized, using examples ranging from expropriation of private citizens by the state, to private prisons containing stateless detainees.

2.5 Drawing the Public/Private Sector Boundary

The distinction between the public sector and private sector flows conceptually from the public/private dichotomy and has been prone to a closely related set of challenges when applied in practice.

2.5.1 *Defining the public sector*

References to forms of economic activity as belonging to the "public sector" or "private sector" date at least to the 1930s(Oxford English Dictionary n.d.). However, the concept of a public sector as a discrete entity has deeper roots in the idea of being a "political" realm (Gavison 1992). The form and function of the public sector has its antecedent in ancient Greece in the *Polis*, a collaborative or cooperative site where ancient Greeks used reason to make decisions on behalf of the state in pursuit of just outcomes (Pitkin 1981). The word political comes from the Greek *politikos*, "of or pertaining to the polis, the city-state." Aristotle describes the state in terms a partnership among the citizens towards the common good. The polis is made up of citizens as a political-ethical community (*koinonia*), a "collection of parts having some functions and interests in common" (Miller 2017). The notion of citizen is limited to those with full rights to participate in deliberative/judicial capacities.

The public sector is sometimes demarcated based on ownership; it is collectively owned by the citizens in a place, as well as managed by them. It is roughly equivalent to the government or the state (Pitkin 1981). Public sector agents are entrusted to act *on behalf of* their administrative office or constituency (Howard 2014). The legitimacy of the state in this capacity is derived from the allegory of the social contract-- that shared interests are most effectively served through cooperative action, or a coercive agent that works on behalf of law and order.

2.5.2 *Defining the private sector*

The private sector is demarcated by its ownership by a specific individual or group. It is also classified on the basis of who benefits from the organization; whereas "commonweal" organizations benefit the public, business organizations benefit their owners (Blau and Scott 1962). Neoclassical economists like Friedman conceive of the private sector as a collection of independent firms, which are able to respond dynamically to change, and among which competition works to improve services even as it reduces costs.[14] Many hybrid organizations sit astride these definitional boundaries, such as the Corporation for Public Broadcasting, a non-profit corporation created and funded by the U.S. government. Edge cases demand a more nuanced set of definitional attributes; some definitions consider funding source (Wamsley and Zald 1973) or whether it is a for-profit or nonprofit organization.

Economists also consider the nature of the "goods" (or commodities) controlled by an entity to determine whether it should be considered public or private. Excludable goods are those which one might be barred from consuming, e.g. a sports stadium; non-excludable goods are those that are not possible to bar to anyone, such as clean air. Goods can also be rivalrous or non-rivalrous-- rivalrous goods decrease as one person consumes them, such as a meal; non-rivalrous goods do not decrease with consumption, such as the music coming out of a speaker. Buchanan (1965) prodded at the definitional boundary of public versus private goods by considering ways in which private clubs cooperatively consume shared resources, such as community swimming pools and hunting clubs. These hybrid forms help make the case for a spectrum between private and public goods, along which pure private goods have one owner, and pure public goods have an infinite number of owners (Whittington 2008; Buchanan 1965). Scholarship from New

[14] Neoclassical economics formalizes seminal assertions made by Adam Smith (1789) in The Wealth of Nations; namely, that the marketplace functions as though guided by an "invisible hand." That is, the social benefits associated with complex trade relations arise from the pursuit of individuals' own self-interest. Neoclassicists assert that the maximum societal benefit results from "market efficiency," an optimum resource allocation. Some economists have extended these insights into normative macroeconomic arguments. For example, Hayek (1945) claimed that the price mechanism helps to facilitate trade and the complex ordering of supply chains without need of centralized state control or coordination. He further argued that markets yield greater freedom and social benefit as competition increases and state intervention decreases. So too, Milton Friedman (1962) campaigned against government intervention in the marketplace, claiming it leads to higher prices and waste. This message, delivered through influential public scholarship (Friedman and Friedman 1990), suffused late 20th century debates about the role of the public sector, and informed government initiatives to relegate services to the free market.

Institutional Economics provides a robust theory explaining these hybrid forms and highlighting the crucial importance of contracts and governance structures in subtending even "private" affairs. [15]

2.5.3 *Shifts in the boundary over time*

One account of the distinction between the public sector and the private sector in the United States describes the state as waxing, then waning over time. On this view, the first half of the twentieth century in the United States, the public sector grew in scale and scope of action. Private sector corruption, monopolies, and stock market negligence fostered a sense that the public sector needed to assume a more central role in regulating the marketplace. It assumed a number of responsibilities to aid in recovery from economic catastrophe, provide a safety net to the poor, re-skill the unemployed, fight in two world wars, and institute a host of progressive reforms like social security. These political developments were also founded on two bodies of economic scholarship;

[15] New Institutional Economics (NIE) builds on, modifies and extends neoclassical theory to reflect empirical observations about how markets function. NIE disputes many of the underlying assumptions in neoclassical economics. Rather than asserting that individuals are rational actors receiving perfect information via the price mechanism, it assumes only a bounded rationality-- economic agents cannot ex ante ascertain optimal solutions due to uncertainty, information problems, and costs which are not captured in neoclassical models.
The notion of transaction costs originated in work by Ronald Coase (1937), whose project was to interrogate why organizations exist, if the classical economic assumption is true that that price is able to organize market interactions. Organizations, he argued, exist because there are costs to using the price mechanism --transaction costs--that organizations help to distribute and defray. The transaction costs mean that prices alone are not enough for actors to arrive at an "efficient" outcome, as defined in neoclassical economics.

Further theorizing the way that market exchange takes place in practice, Williamson centers his analysis not on choice in the abstract, but the concrete details of contracts. He emphasizes the extent to which contracts are "unavoidably incomplete" in nature, given stakeholders' opportunism (i.e. guile and self-interest), as well as incomplete information and thus bounded rationality (Williamson 1975; 2002b). The only way to redress such "gaps, errors and omissions" is to revisit the contracts ex-post, evaluating them for adherence to the agreement, and inserting additional governance controls where needed (Williamson 2002; Whittington 2008).

Whereas neoclassical economists had broad normative claims for how the economy should be organized (i.e. reducing the extent to which the government should intervene in the market), NIE does not prescribe a "one-size-fits-all" approach to economic organization. Rather, it considers the role of institutions and governance in reaching optimally efficient outcomes, and pursues optima on a case-by-case basis in "comparative micro-analytic detail, through the lens of contract" (Whittington 2008, 25). NIE advances a research program to analyze the variance and rationale behind how different transactions are safeguarded, "predicting an economizing purpose for structures of governance" (Whittington 2008, 30).

Economist Douglass North (1990) argued that institutional supports are necessary for markets to reach optimal outcomes. Neoclassical economics elides this fact by presupposing the presence of an institutional framework within which consumers and producers behave. NIE asserts instead that "efficient" outcomes in the marketplace are made possible by arrangements of governance (e.g., contracts) and institutional environments (e.g., polity and judiciary) that define the rules of the game of economic exchange and, in doing so, allow for macroeconomic efficiencies in transaction costs and the production costs of goods in society.

first, Pigou's (2013) idea that market activity alone does not always ensure social utility and is prone to "market failures," defined as the inability of the market to deliver an efficient distribution of goods and services; and Keynes' (2016) work on how governments can smooth out economic shocks through adjusting spending and interest rates (Yergin and Stanislaw 2002; Whittington 2008).

Neoliberalism. In the wake of a widespread recession in the 1970s and 1980s, political leadership attempted to kindle economic growth by applying neoclassical economic arguments advanced by Hayek (1945) and Friedman (1962). These "neoliberal" policies are defined by their interest in removing constraints from markets across all sectors of society by deregulating major industries, loosening protections for organized labor, reducing corporate taxes, and enhancing global capital mobility (Brenner and Theodore 2002). According to Newman and Clarke (2009), accounts of neoliberalism fall into three broad categories: first, as an anti-statist project to change the conditions of capital accumulation to allow for intensifying accretion of resources (c.f. Harvey 2005). A second approach is advanced by Peck and Tickell (2002), who view neoliberalism as a function of the state facilitating capital accumulation. A third strand works from a Foucaultian tradition to examine how government can work "at a distance" by producing subjects who self-regulate. Much contemporary work on changing state-society relations adopts neoliberalism as a lens; Newman and Clarke (2009) caution that this lens may present a totalizing account of neoliberalism as omnipresent and omnipotent, ignoring the multiplicity of factors on the ground. In a similar vein, sociological historians point to the role of the state in fostering private enterprise and depending on volunteerism throughout U.S. history, undercutting the view that neoliberalism is a recent phenomenon (Clemens and Guthrie 2010; Morgan and Orloff 2017).

New public management. This time period also fostered a series of political reforms collectively known as new public management-- especially in the US, UK, and Commonwealth countries. Variously characterized, new public management seeks to apply the market logic of competition, disaggregation, and incentivization to the public sector performance and practices to address government waste (Dunleavy et al. 2006). The public sector sought to adopt private sector management styles, such as product-based teams, top-down management, performance measurement, and leaner budgets (Hood 1991; 1995). Citizens are re-imagined as "clients" or "customers" of government services. Reforms to the bureaucracy were intended to increase the efficiency of service delivery. Governments also came to understand their role as a catalyst for

private enterprise (Osborne and Gaebler 1992). As a result, the state has increasingly ceded ground to the private sector for the provision of services, embracing a technical managerial role over direct service provision (Miraftab 2004).

Governance. As the role of the state diminished in the wake of neoliberal reforms, theories of governance conceptualize private sector action as playing a role in the changing nature of power. This focus grew out of a behaviorist turn in political science scholarship, which focused on the policymaking impact of the broad network of non-state actors, such as advocacy groups and political parties (Bevir 2011). Theories of governance move beyond a "false dilemma between centralized regulation and deregulatory devolution" to embrace new forms of voluntary, cooperative, or delegated participation as forms of power exercised by the private sector (Lobel 2004). Governance studies scholars examine the exercise of power through this lens. For example, in her work on the form and function of administrative law, Jody Freeman (2000) argues that the private sector can be a source of effective regulation. Others have used the lens of governance to critically interrogate public engagement in the adoption of emerging information technologies (Ashton, Weber, and Zook 2017; Burgess 2014).

2.5.4 *Boundary work*

Rather than assume the state boundary as fixed or stable, sociologists advance a theory of its formation as a dialogic process, which is visible at the inception of new programs when still in flux. The government contracts or partners with the private sector to provide a service via "associational policy forms" as hybrids of the two. Such choices are motivated by a desire to reduce costs, draw on enterprise capabilities, overcome political roadblocks, or mitigate concerns over large government (Mayrl and Quinn 2016). In the process of creating these associations, political actors are doing "boundary work" to affirm what the boundaries around the state are; this boundary work has three mechanisms: (i) demarcating the boundary explicitly, (ii) politicizing any changes to resource allocation, and (iii) patterning programs on existing models (Mayrl and Quinn 2016). Sociologists trace the boundaries around the state by detailed empirical accounts of material and other flows.

Instead of committing to an epochal account of how a single boundary shifts over time, Newman and Clarke (2009) use two conceptual lenses for tracing plural, emergent forms as they

are made and remade: articulation and assemblage. *Articulation* is the work of mobilizing discourse (e.g. the meanings of words, strategies of framing) and people to produce a political rationale supportive of a project; *assemblage* is the work of taking heterogeneous "policies, personnel, places, practices, technologies, images, architectures of governance and resources" to make cohesive wholes (Newman and Clarke 2009, 9). Simply put, they adopt ways of looking at public and private that allow for their emergent formation, supporting situated empirical inquiry into how these entities are entangled but come to be solidified and made again.

2.5.5 *Contracting and public-private partnerships (PPPs)*

When provisioning a service, public agencies typically weigh decisions about whether it is more beneficial to provide a service internally or outsource it to a firm. Any service can be broken down into a series of tasks--including design, finance, construction, operation, and maintenance. Agencies can 'make or buy' any of these tasks for a given project. Contracting is motivated by a desire to promote cost savings and efficiency in service delivery. Positive outcomes are contingent on the public sector's capacity to assess and manage providers effectively (T. L. Brown and Potoski 2003). The extra management capacity required to manage these contracts sometimes exceeds the cost savings of outsourcing these services (E. Sclar 2001).

Governments have the ability to choose among a range of contracting options; including (i) operation, maintenance, and service contracts where the infrastructure has already been built but the private sector is believed to manage it more efficiently; (ii) build, operate and invest contracts in which the private sector builds and operates infrastructure to recoup their investment before transferring ownership to the government; and (iii) joint ventures, in which both sectors participate in benefits, costs, and risks (Koppenjan and Enserink 2009).

When a firm provides more than one such task on a project, for example by helping to finance and construct a service, it is known as a 'public private partnership' (PPP) (Whittington 2012). Thus, there is a large set of possible PPP configurations. Some categorize PPPs as a form of privatization (Linder 1999; Burki, Perry, and Dillinger 1999); for instance, Miraftab (2004) calls them a “Trojan horse of neoliberal development.” Others draw a contrast, in that privatization reifies the public-private boundary, whereas PPPs blur it (Starr 1990; Squires, Gottdiener, and Pickvance 2012).

2.5.6 *Infrastructure, information, and software as a service*

Of particular interest for my study is contracting for providing public infrastructure. The term infrastructure can be applied to any networked system that subtends the flow of goods and services to allow the mediation of exchange across distance (Edwards 2003; Graham and Marvin 2001). Infrastructure financing, like other forms of public services, is a battleground among vying political ideologies. Whereas some claim that privatization will help governments afford infrastructure investment given its high initial capital requirements, opponents point to government taxing and bond powers which permit them to afford public infrastructure spending. Previous work has critiqued private sector provision of infrastructure for its limited focus on ensuring equity and sustainability in favor of an interest in maximizing return on investment (Koppenjan and Enserink 2009). Graham and Marvin (2001) caution that market liberalization provides perverse incentives to large firms to service urban elites and cleave away from those who cannot afford it. They call for attention to "distributional justice" by asking who is served by networked structures and who is not.

Information infrastructure. These critiques have assumed particular salience in scholarship about information infrastructures. Previous work notes the intrinsic properties of information markets lend themselves to natural monopolies, including high fixed costs, low marginal costs, and high cost of switching providers (Shapiro and Varian 1998; Varian, Farrell, and Shapiro 2004). When cities outsource data-intensive services, they may therefore be subject to large initial investments and vendor lock-in. Nevertheless, many public agencies to date have been inclined to contract information services to firms based on a sense they are better positioned to leverage innovative technologies for efficient use of resources. Indeed, the focus on efficiencies has meant that smart cities data collection programs tend to be contracted to a single firm instead of multiple smaller providers; in such cases the terms of these vendor agreements plays a key role in delivering high quality services.

Some scholars have critiqued the view that large technology firms can bring new efficiencies to government. Söderstrom, Paasche and Klauser (2014) argue that a few firms' market dominance is the result of successful "corporate storytelling," effectively "channel[ing] urban development strategies through the technological solutions of IT companies" (p. 308) as an

obligatory passage point. Despite critical scholarship on the role of large firms in smart cities programs, dissenting voices emphasize the heterogeneity and variegation of contemporary smart cities programs, "which comprise multinational high-tech firms but [are] not limited to them" (Blomley 2005, 337). McNeil (2015) similarly highlights that corporate vendors like IBM are not all-powerful; he avoids a totalizing account of firms' dominance in the market by a close examination of smart cities programs as incipient, experimental efforts by firms to make markets. Other scholarship points to the need for public agencies to attend carefully to the terms of their agreements with vendors. Whittington et al. (2016) found large gaps in data privacy, security, and accountability within information infrastructure vendor agreements in the City of Seattle and argued in favor of public agencies to hold contractors for data intensive services to high standards in the language of their vendor agreements.

Software as a service. In a 'software as a service' (SaaS) business model, the vendor provides sensors, storage, software, and analytics to the customer organization-- in this case a municipal government. SaaS services are billed in two different ways; either at a fixed "subscription" rate over the term of a contract or based on the cost of the computing services rendered. The adoption of SaaS vendor contracts is in part driven by public agencies' desire to save on fixed costs for IT based services—which becomes more important as budgets decrease. For example, the CEO of open data platform provider Socrata said that his company's SaaS "allows our customers to break from the outdated legacy enterprise [system] gridlock, and results in significant budget relief for government IT departments" (Barns 2016). While saving costs on IT infrastructure is a primary driver of SaaS adoption by organizations; other factors include a desire to gain access to new technology, and to move from a role of deploying and maintaining IT services to managing them (Susarla, Barua, and Whinston 2009).

2.6 Drawing the Public-Private Information Boundary

The public-private dichotomy has an equally strong role to play in the descriptive and normative classification of information. Both public and private information may be defined by their content, their context, or in the nature of the interests associated with them. The definitional boundary between the two is not fixed, in that it often depends on the context of use (c.f. Nissenbaum 2004).

Information stewards must weigh both public and private claims before reaching a situationally specific classification.

2.6.1 *Definitions of private information*

Recall that the conceptual lineage of "private" comes from *oikos*, the household (see Part I). In Ancient Greece, the household was the site of bodily, intimate, and material matters (Arendt 2013). Some definitions of private information resonate with this conceptual lineage, in that they cohere around the "'zone of intimacy"-- marriage, divorce, sexual relations, procreation, child rearing, abortion, and so forth" (Cohen 1997, 140). However, what constitutes private information may not be reducible to its content or genre. Law scholar Robert Post (1989) asserts that it is not possible to ascertain whether information is private based on the content of that information, but rather one must instead "have some notion of the circumstances surrounding the revelation of that information" (p. 957, 980, 981). In a similar vein, philosopher Helen Nissenbaum (2009) also asserts the primacy of context in defining a privacy violation.

The following survey of conceptions of private information draws extensively on privacy scholarship; however, privacy and private information are not coextensive.[16] The term privacy is used broadly to "extend over information, activities, decisions, thoughts, bodies, and communication" (Nissenbaum 2004, 105). Solove (2008) categorizes influential definitions of

[16] Where scholars differ on how private information (qua "privacy") is defined, there is wide acknowledgement of its vital function in society in guiding human behavior. Privacy scholars describe privacy as encompassing different forms; physical/spatial/bodily, decisional, associational, and intellectual. As is the central concern of this chapter--privacy is also widely held to be informational, in that it relates to the state or right of control of information about the self. Alan Westin's (1968) seminal account of privacy describes it as the state of controlling access over personal information and asserts a normative right to do so (Moore 2003). This 'information control' account has strongly influenced contemporary readings of other definitions of privacy; Jean L. Cohen (1997) reads Warren Brandeis' (1890) 'right to be let alone' as essentially informational: "control over the acquisition, possession, and spread of information about oneself, along with control over access or attention by others" (p. 140). Nissenbaum's (2009) definition of privacy is also informational in kind, in that she regards privacy as constitutive of the norms regulating the flow and distribution of information. While emerging information technologies diminish the barriers to certain kinds of personal information (such as DNA, or closed-door conversations), privacy interests still adhere to them (Moore 2003).

Another account of privacy as essentially informational focuses on the privacy harms of information disclosure. In his account of privacy as preserving our ability to relate differently with different people, James Rachels (1975) also mentions four harm that result when information is inappropriately divulged; such disclosures about one's personal life, medical records, or lifestyle can lead be embarrassing, threaten employment status, and diminish access to credit and insurance. In passing, Rachels (1975) also touches on the role of privacy in protecting competitive interests. Although he refers to potential "disadvantage to Bobby Fischer...analyz[ing] a chess game," the same insight holds for commercial competitive interests (p. 1).

privacy into six general, overlapping types: (1) the right to be left alone (Warren and Brandeis 1890), (2) the ability to shield unwanted access to the self by others (Bok 1989), (3) the ability to keep matters secret from others, (4) the ability to control information about the self (Westin 1968), (5) the protection of one's personhood, personality and dignity (Bloustein 1964; Thomson 1975; Prosser 1960), and (6) the control or limited access to one's intimate affairs (Inness 1996).

2.6.2 *Private information as intellectual property and data ownership*

The public/private sector dichotomy is related to the distinction drawn between public and private information. Susan Gal (2002) writes that the relationships between these two domains are not coincidental or inconsistent. While Pitkin (1981) claims that the private sector of "private enterprise" has little in common with the notion of intimate or "private information," Gal (2002) attributes them to the same semiotic logic that yields both meanings; "when analyzed as a 'communicative phenomenon, a product of semiotic processes' it shows a complex and systematic logic that explains this usage" (p. 77). Conceptual resonance between the private sector and private information is evident in intellectual property laws such as trade secret, which allow private companies to shield information central to their competitiveness from public view (see Section 2.6.3.3).

The metaphor of ownership over information has long applied uneasily to information in both analogue and digital forms; information has peculiar properties which enable it to be readily shared. Information is non-rivalrous, meaning that it can be held by more than one person at once. It is non-corporeal, or formless; and non-rivalrous, such that one person having information does not decrease the amount of information held by another (Hettinger 1989). . The tensions in protecting information as property are evident from the spate of challenges in enforcing ownership claims each time new tools to copy information become available. Nevertheless, the law recognizes ownership rights to information in intellectual property law.

Indeed, one account describes privacy rights as a form of personal ownership over information (Samuelson 2000). [17] Moore (2003) examines privacy rights as a possible sub-type of

[17] Defining privacy rights as a subtype of property rights mends an ostensible fissure in privacy interests; those asserted by persons, and those asserted by commercial enterprises. While Pitkin (1981) claims that private enterprise has little to do with the notion of private information, Gal (2002) attributes them to the same conceptual and semiotic framework.

property rights, as both pertain to the control of "physical manifestations or tokens of some abstract idea or type" (p. 218). Defining privacy rights as a subtype of property rights mends a fissure in scholarship about different types of privacy interests-- those asserted by persons and those asserted by commercial enterprises.

2.6.2.1 Data ownership

Some scholars have continued in this vein to argue that individuals have an ownership claim in data about themselves. Data ownership has been explored as a means to acknowledge individual autonomy and dignity in the governance of sensitive information, such as medical records (Hall 2009). More recently, data ownership has gained increasing salience as a policy proposal to fight online power asymmetries with platforms. Prainsack (2019) dichotomizes data governance strategies into "individual control," which includes but is not limited to advocacy for data to be individual property protected by property rights, and "collective control" approaches, which includes advocacy for data to be protected and used within commons models. Property rights are described as a set or bundle, including the right to enjoy, dispose, or exclude others from the use of property.

Law scholars Paula Samuelson (2000) and Paul Schwartz (2004) both argued that a data ownership regime would obligate firms to compensate persons for the use of their personal data. Samuelson (2000) describes the affordances of a property-based approach to data protection as granting individuals the ability to recapture some of the value of data about themselves; and limited in that it differs so much from existing regimes of treating information as property (i.e. intellectual property law) as to both weaken existing IP protections for creative works and undermine the case for personal information as in need of protection. Both Samuelson (2000) and Schwartz (2004) have both explored data ownership as a basis on which firms would have an obligation to compensate persons for the use of their personal data.

2.6.3 *Definitions of public information*

Scholarship on publicity and public information has developed apart from discourses on private information and can refer information that is descriptively visible to all, normative transparency

obligations of the state, or information that has value for society as a whole.[18] Normatively, the question of what information should be made public is more complex. Public information is less explicitly theorized than private information per se; it is generally considered in terms of its constituent elements or subtypes. Notably, legal scholar Woodrow Hartzog (2019) describes the concept of public information as a popular fallacy, without fixed definition in either law or policy; he asserts that the political consequences of the appellation demand that the term must be more clearly conceptualized. Harzog points out that to date, the term "public information" is used three ways: descriptively; that is, when it is freely accessible or widely known; negatively, that is, when it is not withheld as private; or by designation, as information that controlled or released by state actors. Previous work resonates with Hartzog's typology. I present them here as (i) the public right to know (such as freedom of information, and sunshine laws), (ii) the distinct, yet related "open" government and transparency discourse, (iii) other communal interests in information, such as public health, and (iv) public domain and related public interests in intellectual property.

2.6.3.1 Government transparency

The public "right to know" is a widely-held human value, yet the "precise reasons [for it,] and the extent to which publicity should be guaranteed are not straightforward" (Gosseries and Parr 2017). Moral philosophers offer different accounts of what information should be public, and why (c.f. Mill 1869; Meiklejohn 1948; Volokh 2000). Here, I focus primarily on accounts that deal with the right of the general public to government information. While Kant (1915) asserted the importance of political maxims to be defended publicly in his 1795 "hypothetical publicity test," utilitarian approaches tolerate governments' use of secrecy in exigent circumstances (Gosseries and Parr 2017).

Government transparency norms have taken form in law and policy. In the United States, the flow of public information was intended to educate the populace to support their self-determination and government. As James Madison writes to W. T. Barry (August 4, 1822):

[18] The etymology of the word 'public' has several denotations; 'official,' 'representing the community,' 'belonging to the people as a whole' or something open or generally known (OED public definition). So too does the notion of public information sit astride these two facets-- pertaining to the state and representation, as well as to information that is available to the community in general. Descriptively, public information is that which is viewable or knowable by people in general, or open in plain view. This expectation is also reflected in American law as the "plain view doctrine."

> *"A popular government, without popular information or the means of acquiring it, is but a prologue to a farce or a tragedy; or perhaps both. Knowledge will forever govern ignorance. And a people who mean to be their own governors, must arm themselves with the power knowledge gives"*

Emerson (1976) reads the First Amendment to the U.S. Constitution as inclusive of such a 'right to know.' He reads freedom of expression as having two parts: the right to consume information, and the right to obtain it for the purpose of transmitting it to others. Emerson argues that the 'right' to know be recognized as such independently of the right to freedom of speech, in that it focuses on the "affirmative aspects of the first amendment," rather than a negative right to be free from state interference (p. 2).

In interpreting the First Amendment, The U.S. Supreme Court has recognized a right to know in Lamont v. Postmaster General (1965) in which citizens can receive 'foreign communist propaganda' without notifying the government (Emerson 1976). In a 1969 case about consuming pornography in the home, Stanley v. Georgia, the decision read, "It is now well established that the constitution protects the right to receive information and ideas." However, the Supreme Court has on occasion decided otherwise.

The federal government has also passed laws that uphold a public right to know. In 1966, Johnson amended the information access part of the Administrative Procedure Act (APA). These amendments, widely known as the "Freedom of Information Act," (FOIA) (U.S.C. § 552). FOIA was originally tied to a section of the APA that gave government agencies the ability to exercise wide discretion over which of its records it would choose to publish. There are nine exemptions specifying types of information that can be withheld from disclosure. These exemptions span interests in national defense, internal personnel files, trade secrets, invasions of privacy, privileged communications, law enforcement proceedings, and any applicable statutes.

In the event that a public records request is denied, the agency may have a duty to provide its rationale-- be it to protect privacy, public safety, or a trade secret. In some cases, providing a precise reason for why information is being withheld from disclosure may have an indexical relationship to the information itself. In other cases, even confirming the existence of a particular record with an explicit denial to disclose it would reveal information about the document's content; this is known in the US as a “Glomar response.”

2.6.3.2 Open government discourse.

The term open government carries a different cultural valence than the “right to know” discourse that preceded it. Scholars generally agree that the term has been influenced by the “open source” vs. Free software debates of the 1980s and 1990s. Whereas Free software licenses collaborative work with no restrictions on distribution, modification, or use, Open Source development products are proprietary. Nathaniel Tkacz (2012) traces the roots of the 'open government' movement to such an open source rationale, finding that open government is similarly "business-backed" and "compatible with a new form of capitalist accumulation" (p. 393, 395).

Others disagree that the open government movement was borne out of an open source rationale. Yu and Robinson (2012) observe that the term open government originally carried the same “hard political edge” of politically sensitive disclosures of the 1950s and 1960s; it was only with time that the term became conflated with that of open technology (p. 178).[19] Yu and Robinson critique this development, claiming that it led to a blurred and empty conceptualization of openness in government, in which government might be called “open” for dedicating resources to publishing data instead of to meaningful transparency reforms.

2.6.3.3 Collective interests in information

Another way of characterizing public information can be garnered from normative arguments for the communal or collective interests in some types of information. These debates are usually framed in terms of a need to balance privacy with other human values in context. For example, functioning of the markets is made most efficient by public access to information (Morris and Shin 2002). Posner (1981) remarks that there is a social cost to privacy in terms of the slowed efficiency of decision-making. He uses the example of employers in the labor market looking for a reliable employee; he argues that some information about the employee's mental health and other impediments are first salient to the greater good of the employer and the firm. Second, there is a public health rationale for sharing some types of private medical information. If privacy in public health data can risk others' health and safety or diminish available data for research and evidence-

[19] Note that the strong pro-transparency commitments of the Washington State Public Records Act, passed by ballot initiative in 1972, attest to this aspect of Yu and Robinson’s argument.

based medicine, it follows that there is a collective or 'public' interest in some types of private information.

2.6.3.4 Public domain and other public interests in tension with intellectual property

In addition to the exceptions covered above, limits on intellectual property laws also help to trace the contours of public information. In this regard, public information is defined in relief to that which can be covered under the U.S. intellectual property regime. There are four types of intellectual property under U.S. law; copyright, patent, trade secret, and trademark; of which the first two indicate a public interest in some types of information. Limitations on copyright recognize a *public interest* in educational, interpretative, and other uses of otherwise protected material. The U.S. Patent laws are also written in such a way as to take the public interest in proprietary information into account. In order to hold patent rights, the creator must provide detailed information about the invention to the U.S. Patent and Trademark Office (USPTO, which must evaluate whether the invention qualifies for protection. If the patent is offered, the USPTO publishes the schematic provided by the inventor for the benefit of the general public. Publicizing this information is intended to accelerate the pace of technological innovation in the U.S., while incentivizing inventors to continue to create. Exceptions to legal claims of trade secret are determined by courts on a case by case basis; these exceptions also recognize public interest in even trade secrets to be disclosed on occasion.

2.7 CONCLUSION

The implications of this literature review for my research are to highlight the political terrain on which my work takes place. In particular, I focus on data crossing the passage points between the public and private sectors in contracting relationships. Theory in sociology and organizational studies considers contracting relationships as "associational," that is, an active staging ground where state boundaries of the state are made and re-made through resource flows (Mayrl and Quinn 2016). Recall that Gal (2002) reads the terms public and private as always inherently normative. From this departure point, my research examines claims to or against data access as normative rationales about data as public or private interests, with a focus on the accountability and ownership arguments raised to support these claims. Given that the classification of information as either

public or private is as fraught at the dichotomy itself, my work describes these distinctions as they are applied and contested by multiple actors.

Chapter 3. BACKGROUND AND METHODS

In this chapter I provide background as to how I came to work with local public agencies on problems related to data sharing and governance and the salience of the Washington Public Records Act for local transparency, privacy, and data sharing concerns. I describe how my participant observation on a university-based data sharing intermediary effort with Prof. Jan Whittington exposed me to the broader context for this study as a collective action problem, and in turn informed my case selection for this work. While the confidential nature of the discussions behind the Transportation Data Collaborative necessarily limited the information I could include from that work in this study, the centrality of my work on that project to this dissertation gave me access to the confidential perspectives of key stakeholders and in turn shaped the questions posed in this study. I go on to describe my methods as ethnographic, taking cues from the methodological focus on breakdown in my focus on conflict and failure (Star 1999) and van Dijck's (2013) approach to political economy in my attention to law, policy, and contractual agreements. In my analysis, I draw on Mol's (2003) approach to tracing concepts by inverting them, examining the concrete, situated practices that constitute their "enactment" in practice.

3.1 BACKGROUND

Seattle, Washington is a city in the Northwest United States with a long history of industrial change, technological innovation, and activism. A city once reshaped by the western expansion and the Gold Rush, by 2010 Seattle was witnessing another rapid influx of labor from the

technology industry, putting pressure on housing and the cost of living. Although the pace of change was accelerating, the city was experiencing a continuation of prior trends; the largest employers in the region, Microsoft, Amazon, and Boeing, had long attracted and helped to foster a large pool of local engineering talent. The 40-year legacy of engineering in the region had shaped city governance and policy discussions in turn; residents lent their technological expertise to government service, or through local non-profit, advocacy, and activist organizations. Technology policy debates were roiling when I moved to the city in 2013. Here, I provide the broader context that motivated my work, my access to the field, and rationale for selecting the cases I chose to report on. I also explain how my role as a participant observer personally enrolled me in many of the same dynamics of information disclosure and secrecy that I explore in my work.

3.2 Municipal Data Privacy in Seattle

In 2013 Seattle was the site of rising scrutiny of government use of data collection hardware and software. This scrutiny preceded the Edward Snowden revelations about government surveillance yet intensified in their wake. Local activists had organized in response to news reports that the Seattle Police Department had acquired surveillance technologies without clear policy guidance or oversight, such as a drone aircraft, a closed-circuit camera network, and mesh network infrastructure (Crump 2016; Young, Katell, and Krafft 2019). Organizing around these technologies resulted in concrete policy changes, like the 2013 Seattle Surveillance Ordinance, as well as new activist groups like the Seattle Privacy Coalition, which joined other local groups like ACLU Washington to advocate for stronger protections for Seattle residents' privacy more broadly. I waded into this space slowly, first attending then co-organizing with Seattle Tech Activism Third Mondays, a group that convened local privacy advocates with crypto evangelists in a worker-owned coffee shop in the Capitol Hill neighborhood (and later, the office suite where The Tor Project anonymity-preserving browser was headquartered) to share skills and knowledge.

This fomenting local interest in technology policy also found expression in new opportunities for research; a local (and simultaneously transnational) firm joined foundations in granting seminal funds for the University of Washington Tech Policy Lab, co-founded by Professors Ryan Calo, Batya Friedman, and Tadayoshi Kohno in 2013. By 2015, the City of Seattle and Microsoft funded the Lab to explore the implications of municipal data collection on privacy,

accountability, and public trust in a project that would be my first formal research foray into local government. The City was in the process of architecting a new data privacy program, informed in part by an expert advisory board and by our study (i.e. Whittington et al. 2016). I was hired by Ryan Calo to conduct focus groups with a range of local stakeholders to elicit their hopes and concerns for how municipal government handled identifiable data. I spoke with people in local industry, city employees, Seattle residents, civic hackers, and privacy activists. Co-PI Jan Whittington also invited me to participate in interviews with personnel responsible for data-intensive programs in the city, and how the resultant data was handled under the state public records laws. The initial goal of our work was to explore municipal data as it is made "open" be it via hack, open data, or public records disclosure. Our research came to suggest interesting frictions between the high-transparency intent of the state freedom of information law as written, and the new volumes and types of data being collected under a "smart cities" paradigm of data-driven innovation. This set of frictions ultimately became my central interest.

3.3 The Problem: Privacy and Public Records in Washington State

The Washington State Public Records Act (PRA, RCW 42.56) was passed by ballot initiative in 1972. It is one of the most pro-transparency state laws of its kind, mandating that WA public agencies must disclose most records they prepare, own, use or retain. Antecedent initiatives to the PRA were originally motivated by a push for campaign finance transparency but in the language of the state, encompass most any government records available for disclosure, such as letters, contracts, resolutions, electronic records, emails, photographs, videos, MP3s, databases, voicemails, and instant messages. In the absence of a specific exemption, agencies must disclose requested records—and are subject to per-day and per-record fines for non-compliance with the Act.

The law was described to me by public employees as being "high stakes" and stressful for public agencies, given that courts may fine agencies who fail to comply in amounts accruing by day and by number of pages. A 2017 report from the State of Washington Auditor's Office found that compliance with the law costs state and local governments a collective total of $60 million dollars per year (Washington State Auditor's Office 2016). Many local government employees that I spoke to in Washington State were very concerned that the law had not been amended to

give public agencies more discretion. They attributed its enduring strength the dedication of a strong pro-transparency constituency and a formal lobbying group, the Washington Coalition for Open Government, as well as formal oversight and monitoring conducted by the state legislature's Sunshine Committee, a body which reviews exemptions to the Act as they accrue in statutes over time. Others highlighted that the state legislature is exempt from the PRA, shielding legislators from what local government critics perceive to be the law's privacy harms.

In previous work, we found that the public employees, local residents, and activists were concerned about the Washington Public Records Act (Whittington et al. 2016). I conducted focus groups with city employees who reported that they felt personally exposed by the kinds of personal information publicly available about them, such as their home addresses, income, marital status, and other sensitive information (such as membership in the LGBT group for city employees). They were also sensitized to the privacy implications of the Public Records Act for residents; a woman in the Parks and Recreation department reported her discomfort at the granularity of data collected by her department and which she felt legally required to provide on request; the potential for such data to be misused gave her pause.

In contrast, privacy and civil liberties advocates in my focus groups were concerned about the discretion that employees had to withhold information under the Act, and worried that any exemptions would subvert the aims of government transparency. Members of the general public reported the harms that they had experienced with frictionless, unaccountable public records proliferation from outdated background check results to microtargeted mortgage rates. And civic hackers reported that they would prefer records be posted as open data rather than relying on public records requests, they pointed to the city's vendor for the open data portal as a new intermediary, interfacing with the public instead of civil servants.

Conversations Prof. Whittington and I had with employees in the Seattle Department of Transportation attuned us to other potential privacy risks, such as how detailed longitudinal location trace data would be treated under the Act, if ever it were to be requested, and what potential these records held for re-identification or misuse. While at that point the courts had not ruled on that precise question, others had successfully petitioned for similar such records, such as license plate reader data (Newell 2013) and unredacted body camera footage (Conti-Cook 2017).[20]

[20] Washington State court precedent helps to establish how the law should be applied in cases where records are not held by the government directly, such as in the case of government vendors. In 1999, an influential decision on

During the same time period, public agencies across the state of Washington were being rocked by an anonymous records requestor, later identified as Tim Clemens, who had created automated bots to submit incessant and blanket requests for records. At the time he initiated his requests, frustrated and concerned employees in public agencies were attempting to negotiate down the requests by pointing out the infeasibility of fulfilling them, and remarking on the absence of legal means to prevent further automated requests. Clemans was particularly interested in police body-worn camera footage and successfully obtained hours of video from the Seattle Police Department, posting it unredacted on YouTube. Clemans' actions illuminated advocates' concerns with the privacy ramifications of police body camera use and catalyzed what would become a substantive set of reforms in 201to the Public Records Act that enacted new cost recovery measures on certain requests and discarded automated requests (Collins 2017). What at the time seemed to be an uncontained conflict between privacy and transparency for new data types set the backdrop for what would become my fascination with the Public Records Act, establishing it as a context, central concern, and tool for data collection in this study.

A further source of frustration was the way that the Public Records Act was cited as a factor in stalled cross-sector initiatives (as in a partnership between King County Metro and Lyft in Chapter 5), or specific oversight functions (such as when the City of Seattle required rideshare firms to share data to receive a permit to operate). Employees of both firms and public agencies cite the state's public records law as a major barrier to these negotiations; as the high likelihood of disclosure of information shared with public agencies became more clear over time, the same

---

campaign finance disclosures (Telford v. Thurston County Board of Commissioners, 95 Wn.App. 149, 974 P.2d 886) yielded a heuristic device known as the "Telford test" to establish whether a vendor is the functional equivalent of a public agency (Ware 2009). The Telford test is applied on a case-by-case basis and has four parts: (i) whether the entity performs a governmental function, (ii) the level of government funding, (iii) the extent of government involvement or regulation, and (iv) whether the entity was created by government.
Later, the Telford test was used in a case about whether the records owned by a contractor should be released via a public records request. In Cedar Grove Composting v. City of Marysville, 188 Wn. App. 695, the court found that the city's contractor was subject to the PRA., even when the agency initially asserted that these records belonged to the vendor (Municipal Research and Services Center n.d.; Levan 2015).
However, neither the statute of the Public Records Act nor case law provides definitive guidance as to how the law applies to data services provided by third parties, which may present additional complexity. For instance, if data is generated in relational databases on a government's behalf, but only a report of the data is shared with government, which version of the data is a public record? In previous work, Whittington et al. (2016) and Gaughan (2016) both found examples[20] wherein the City of Seattle relied on its vendors to process "raw" data and deliver an aggregated data product (c.f. Kroman 2015). Later, 2018 guidance provided by the City of Seattle (perhaps in part in response to our work) clarified that there is not possible to "evade the PRA by contracting out government functions" (Nadleman and Franklin 2018). This 2018 guidance was consistent with 2017 interviews I conducted on the subject with public records disclosure officers and experts. [20]

agencies were less likely to be able to directly access firm data in the absence of a legal mandate. The potential for disclosure thus disposes firms to consider data sharing with the public sector as risky -- further entrenching a siloed data sharing ecosystem. As a result, although the state's high-transparency public records law aimed to improve public access to data and information, it had the unintended consequence of limiting government visibility into third party systems by encouraging companies to proactively invoke legal mechanisms to withhold any and all data that may be shared with government. This unusual context supported our intervention to create a privacy-preserving data sharing intermediary, the Transportation Data Collaborative, which was my primary vantage point for this research.

3.4 The Transportation Data Collaborative as Intervention

The Transportation Data Collaborative effort began in 2016 as an exploratory project to intervene in the privacy problems inherent in a high-exposure legal context in which sensitive data might be disclosed. Its aim was to explore how to protect data in the transportation domain given recent scholarship that pointed to its sensitivity from a privacy perspective. It drew inspiration from medical data repositories housed at universities to explore how the University of Washington could provide a data intermediary to simultaneously protect research data from misuse while also making it more widely available to researchers. Whittington drew an analogy between transportation data and medical data, saying that the sensitivity and privacy risks of medical records are well known, but that location data could be present similar privacy risk (de Montjoye et al. 2013) but does not have a law like HIPAA to protect individual data subjects. Thinking about data privacy as a collective action problem, she set out to explore the role that a trusted data sharing intermediary could play in data protection. Bill Howe in the UW Information School joined the project as part of his research interest in emerging privacy-preservation and bias removal techniques, and Mark Hallenbeck in the Washington State Transportation Center joined in his capacity not only as a transportation researcher who would use data from the Collaborative in his analyses, but as a close partner with state and local transit agencies and Departments of Transportation whose projects had stalled due to data sharing frictions. In late 2016, I was hired to join the Transportation Data Collaborative initiative to explore the ways that the University of Washington currently protects

data, beginning with interviews with personnel and its medical data repositories, and to explore whether the university could create an analogue in the transportation domain.

My original intent was to write a case study about the Transportation Data Collaborative as part of my research, but the sensitivity of these discussions over the two-year effort foreclosed the possibility of informed consent. The initial funder for the project was a large multinational corporation interested in how the repository could be a proof of concept for their cloud projects to be used in secure multi-party computing and encryption; this strand of the work was protected by a non-disclosure agreement that the principal investigators signed on behalf of our team. As the project matured, it came to encompass a series of conversation both within the university administration and beyond it, including with major rideshare companies, vendors to transportation agencies (such as those producing sensor networks used in transportation planning), and agencies themselves. Many of these conversations were confidential; conversations within the university to assess the risk of ingesting location data were protected by attorney client privilege, and our team granted major firms' requests to sign non-disclosure agreements in advance of key discussions. My role on the project encompassed a range of research, project management, and strategic functions that enrolled me in these private discussions. Without a discrete part of the project for which I could readily subject participants to informed consent, this work instead came to shape my thinking about the problems I present in this work.

The Transportation Data Collaborative ultimately yielded partial success in creating a usable model and series of legal agreements that had been thoroughly vetted and approved by all major stakeholders within the university as well as the deans of the three Co-PIs schools. Over more than 18 months, Prof. Whittington and I achieved this milestone through meetings with the Human Subjects Division, Office of Public Records and Open Meetings, Office of the Chief Information Security Officer, Chief Privacy Officer, Office of Finance, Office of Research, Information Technology, Office of Research, Comotion (for technology transfer), and the Washington State Attorney General (the University's lawyers). I was responsible for drafting key policy documents responsive to both university and state law requirements as defined in these meetings, including an IRB Protocol, a Security Plan, a Privacy Policy, a Records Retention Schedule, and a Public Records Request Response Plan. Jan and I received legal assistance on other key agreements that I co-authored, including the Data Sharing Agreement, Data Use Agreement, Affiliate Agreement, and attachments such as membership tiers and pricing.

Beyond the university, we were also taking at meeting twice a week over two years (2017-2018) with key public and private stakeholders to discuss the initiative and how our project could help to address their data sharing concerns. Firms we met with discussed their data sharing concerns. Many public agencies we met with were initially skeptical about a data sharing intermediary. In their view, much of the data in question belonged to the public or was possible to protect with anonymization techniques. As part of the "pitch" to agencies in support of using the Collaborative as a data sharing intermediary, our team shared resources about privacy risks in location data, the limitations of anonymization, and the potential for the university to serve in a supportive role. As part of these pitches and "lead development," I co-authored a white paper on data privacy, slide decks, proposals, and other materials reflecting what we were learning from stakeholders about their needs for the Data Collaborative, and how that would be reflected in the legal-technical design of the repository. Not including those firms with which we signed non-disclosure agreements, these meetings included Challenge Seattle, Acyclica, INRIX, Siemens, Amazon, Lime Bike, Spin, Ofo, BMW, Los Angeles County Metro, Sound Transit, King County Metro, the City of Bellevue, and multiple departments in the City of Seattle.

These conversations, as needs assessments with key stakeholders, illuminated Whittington's characterization of data privacy as a collective action problem, and my increasing awareness of the irony of our supporting firms' trade secrets in order to make more research data access possible. We came to understand the Collaborative as responding to a series of competing interests: individual residents are concerned for their privacy, firms about their competitiveness and how to mediate multiple incoming data requests from government. Researchers are interested in doing essential basic and applied analysis with data. The government requires data to provide oversight and ensure equity in service provision, and the public wanted both accountability over government and to see the creation of innovative applications that could address complex regional problems, like traffic congestion. All of these stakeholders also cited the high-transparency context of the WA Public Records Act as a key impediment to data sharing. Within the University, we had identified an exemption in the statute that would support safer data sharing as long as it was being used for research.[21] Based in part on our longer-term engagement with the City of Seattle on data

[21] Specifically, our legal analysis led us to rely on a specific exemption to the Public Records Act, RCW 42.56.270, for "Financial, commercial, and proprietary information." It reads that "Valuable formulae, designs, drawings, computer source code or object code, and research data obtained by any agency within five years of the request for disclosure when disclosure would produce private gain and public loss" would be exempt, which we interpreted to

privacy, the City was interested in moving swiftly to try out the Transportation Data Collaborative as a partner for data sharing in their bike share pilot.

The bike share pilot became a key proof of concept for the Transportation Data Collaborative as an intervention to address multiple parties' data sharing needs. In the summer of 2017, the City of Seattle began a pilot program for "dockless" bikes, issuing permits for three different companies: Lime, Ofo, and Spin. For the bikeshare pilot with three firms, the city of Seattle provided an important mechanism in incentivizing firms to share data with the Collaborative; they made the permit for the license to operate bike share bikes contingent on data sharing. The firms were given the choice to either the firms could share data with the City of Seattle directly, or with the Collaborative; all three firms opted to share with the university-based Collaborative. The TDC was able to get access to three firms' data over more than 18 months. Initially, our team used the deployment to answer the City of Seattle's questions about the outcomes of its bikeshare pilot. The corpus of data produced during the pilot has provided research data access for work on transportation equity and bias in Bill Howe's lab.

My firsthand experience helping to build the Transportation Data Collaborative as a university-based data sharing initiative also attuned me to the value tensions at stake in securing access to sensitive identifiable transportation data. Through drafting policy and protocols for this effort, I became familiar with the challenges that public agencies had in accessing data from their partners and vendors. As a result, both this work reflects my personal experiences in navigating trade secret and other confidential information at various scales in the case selection, data collection, and presentation of this work.

3.5 Motivations and Reflexivity

My positionality in the field site was deeply shaped by the access I had been given by my work with public employees and related stakeholders like civic hackers and tech activists since 2014. While my relationships to different participants tied me to the field in multiple ways, there are two that bear mentioning here. First, as a Research Assistant on the Transportation Data Collaborative

support our bid to collect and protect granular data for research purposes (see full language of exemption in Appendix E.)

I was dedicating a significant time to arguing that identifiable location data would be too sensitive for agencies to collect and manage directly, because it risked exposure under our state's high-transparency legal context. As a team, our effort was hoping to convince public agencies that direct access to data would not be necessary in light of the privacy risk of data spillage and promise of finding new data assets within their (indirect) reach. Second, as a Research Assistant on the 2015 work with the City of Seattle and someone who had continued interviewing local activists, advocates, and public employees about the Public Records Act, I felt at the time (and still do) that the Public Records Act posed a large risk to individual privacy and that it could potentially be interpreted in a way that would result in privacy harm. The longer I worked on the Transportation Data Collaborative, the more these two factors merged. I was inclined ultimately to pursue the case studies I report on here discretely from this other field exposure, so as to more easily choose what to include.

Ultimately, I experienced both of these ties to the field in ways that directly enrolled me into the secrecy and intellectual property concerns I report on here as part of the trade secret-keeping apparatus. As previously mentioned, many of these conversations were confidential, covered by non-disclosure agreements signed by lead personnel, or pointed out to me personally by our attorney as subject to attorney-client privilege—which I might waive if not for the ways they include common legal arguments relied on at the university. Hundreds of pages of meeting notes were not possible to report from. Nevertheless, these encounters also provided me with unique access to each party's candid interests and views, which in turn shaped the cases and questions I chose to pursue. Negotiating these sensitive matters in my fieldwork ultimately attuned me to the stakeholder positions and supported me to characterize them in more depth in my case studies.

3.6 Research Approach

In order to examine how data ownership is enacted in practice, I conducted two case studies of contested data access and sharing in Seattle, Washington over 2015-2019 as part of a larger exploratory study of public agency access to contractor data. I adopted an ethnographic approach in order to understand multiple actors' perspectives on data sharing and access frictions in their situated context, and to elicit the social, historical, and technical context for current challenges.

Ethnographic methods afford the ability to produce ‘‘emic’’ (i.e. grounded within participant views and rationales) accounts of phenomena, supporting an analysis of each case on its own terms.

	Public information	Private information
Public sector	Records disclosed	Available to be withheld from disclosure in the event of third-party injunction
Private sector	Confidential vendor data released	Trade secret

Table 4 A matrix mapping possible outcomes from the intersection of public sector vs private sector and public information vs private information

I draw inspiration from ethnographic researchers like Susan Gal and Annemarie Mol who rather than defining a concept seek to explore its multiplicity in situated use. Gal (2002) explores the contradictions at play in the ways the public/private distinction is used in everyday life. Mol (2002) advances an approach to studying a seemingly stable concept (here, atherosclerosis) to show how it is “enacted,” that is, constructed and understood, by different situated actors, such as doctors, patients, technicians, revealing the concept in its multiplicity. In this respect, rather than advancing my own definition of data ownership, I am interested in the ways that it operates on the ground, its contradictions, and how it is deployed by various stakeholders.

3.7 Case Selection as Studies in Breakdown

For this work, I report on two cases of public-private information sharing. Case 1 reports on a public agency and their vendor. Case 2 reports on a public-private partnership between an agency and a firm. The Transportation Data Collaborative initiative played an important role in affording me access to the field, which guided my selection of the particular firms and agencies I chose to examine more closely. In particular, these meetings helped me hear the “off the record” versions of events, including the frustrations and delays, which favored the cases I chose as useful precisely because of their reported problems. Nevertheless, these cases were not unrepresentative of other data access and sharing frictions within the state, region, and nationally; data access disagreements and breakdown led the data sharing intermediary model we piloted at the University of Washington

to receive national attention at practitioner conferences (like the one I attend in Chapter 5 to present on the Collaborative) and to be closely emulated by the Society of Automotive Engineers' 2019 Mobility Data Collaborative.

Identifying contested data sharing relationships was a primary selection criterion for this study. In this regard I move methodologically from Susan Leigh Star's ethnography of infrastructure, which prescribes case studies of breakdown as opportune moments to illuminate their inner functioning. Infrastructure studies shares common epistemological and methodological approaches with ethnographic methods. Star provides methodological notes on using ethnography to study information infrastructure (Star 1999). Her work looks for the "master narrative" in information systems-- the elements of their design that reflect normative assumptions about who the technology is designed for, and for what purpose. She points out that information systems do not usually work seamlessly. They leave gaps that end-users must patch, in a process called 'articulation work.' In general, this work might be done by less valorized workers, like secretaries.

According to Star, information infrastructure can be "read" in multiple ways—whether the researcher foregrounds its materiality, its status as a transactional record, or its status as a veridical account of the world. Tracing to these early studies, common interests in infrastructure studies as a field include rupture, maintenance, and repair of large technical systems, making visible forms of labor that enable them to work and scale, their social shaping, and a critical examination of the interests such systems serve. In part, infrastructure is bracketed because it is hard to access, or embedded in code. It is also background instead of foreground. Star first noticed that infrastructure can be when an eScience system that she helped to design based on three years of fieldwork had failed when it failed to account for incompatible platforms, a lack of resources, and insufficient computing centers—the infrastructure of her field site.

To address this oversight, Star and colleagues offer tricks of the trade to those studying infrastructure ethnographically. Treating infrastructure ethnographically requires a movement that Bowker calls 'infrastructural inversion;' tracing the functioning of the system-- what is usually invisible to users-- by focusing on the system's work, maintenance, and breakdown. Star also notes that different aspects of information infrastructure can be 'read' in multiple ways—whether the researcher foregrounds its materiality, its status as a transactional record, or its status as a veridical account of the world. Star recommends that researchers treat these different representations of technology separately, as they can be hard to disentangle in an analysis.

3.8 Rationale for Ethnographic Methods

Ethnographic work is well situated to capture rich and detailed accounts of practice in light of multiple contextual factors, like system design, laws, policies, history, and beliefs. Ethnographic methods depend on participant observation, structured, and semi-structured interviews, which will allow the researcher to understand variation across informant perspectives. Because the work unfolds over a long period of time, it is able to follow shifting narratives and events over time. Variants of ethnographic work adapted to the study of sociotechnical systems supply a range of methodological techniques to account for the role technologies play in shaping phenomena. Ethnographic methods are also valuable as exploratory methods that can produce new, inductive understandings of phenomena, not over-determined by existing theorizing. It is also flexible enough to be combined with other methods towards a common goal.

3.8.1 *Using ethnography to study networked entities and assemblages*

Ethnography has historically also been well-suited to studying the impact of a technology in a particular place; "our disciplinary bent is to examine the influence of a road in this part of Peru or that part of Niger rather than to analyze the road building as a network" (Larkin 2013). Traditionally, it is not accustomed to considering how networked infrastructures unite a space, or assemble heterogeneous actors such as "politicians, technocrats, economists, engineers, and road builders, as well as road users" (Ibid). Infrastructure has most often sunk below the level of awareness in field studies. "Much of the ethnographic study of information systems [only] implicitly involves the study of infrastructure.... However, it is easy to stay within the traditional purview of field studies: talk, community, identity, and group processes, as now mediated by information technology" (Star 1999).

3.9 Concepts as articulated in and through practices

My method and analysis of how data ownership works is inspired by medical anthropologist Annemarie Mol (2002) whose application of actor network theory to atherosclerosis

begins without a fixed definition. Instead, through situated observations with doctors, patients, families, and medical administrators, she explores how the disease "exist as articulations of the practices that produce them" (Martin et al. 2018). Specifically, instead of beginning with an a priori sense of the concept, she traces how they are performed or "done" anew in different contexts, through a detailed evaluation of the practices by which it is diagnosed. She calls the performance of a disease its "enactment:"

> *"The ethnographic study of practices does not search for knowledge in subjects who have it in their minds and may talk about it. Instead, it locates knowledge primarily in activities, events, buildings, instruments, procedures and so on" (Mol 2002).*

Put another way, by focusing on hospital practices, Mol argues for an epistemological shift from pre-received notions of what atherosclerosis *is*, to a context-driven analysis of the practices from which atherosclerosis *results* as a diagnosis. I adopt this tool to analyze the common rationale of "data ownership," which arose in each of my case studies as an argument advanced in favor or denial of access to data. I use Mol's approach to conceptual inversion as a tool in my own discussion of "data ownership" to describe its conceptual contours, how it is deployed, and how it functions in practice (see Chapter 6).

3.9.1 *A political economic approach*

My work adopts van Dijck's approach to the study of platforms to combine an actor-network approach with a political economic lens. Political economic approaches use organizational structures as the unit of analysis (van Dijck 2013). As it applies to information science, it provides tools for thinking about how institutional actors, such as information technology companies, exert power over users by owning networks (Castells 2013). Whereas classical ethnographic methods focus largely on atomistic accounts of people and practices on the ground (interpreted literally), more recent traditions of ethnographic methods are able to consider the impact of policies, contracts, and other governing factors. Van Dijck (2013) prescribes an approach that builds on Castells' institutional analysis, to also include the technologies themselves, and the content that passes through the network. She frames her approach as looking at six facets of a phenomena:

- Technology – The software and services that turn social behavior into computational architectures.
- Users/Usage – People who participate, produce, consume, and receive a sociotechnical system; van Dijck focuses especially on user agency (or the lack thereof).
- Content – The content a platform contains varies widely with the platform, but may include text, photo, video, or other media that "flow through the ecosystem's arteries" (van Dijck 2013, 35).
- Ownership – The entity that owns the platform, which may by a non-profit organization, corporate firm, partnership, or other organizational form. The ownership model is a core part of its system of production.
- Governance – The mechanisms by which data and traffic are managed; it may include protocols, end user license agreements, terms of service, or in my study, vendor agreements.
- Business models – The ways that the platform generates income.

All six of these aspects of sociotechnical systems were salient in my case studies. In the following chapters, users are public agencies; platforms are the firms that provide proprietary data collection infrastructures; technology is the set of hardware, software, protocols, and data warehousing that produces and manages data; governance is the law, policy, and vendor agreements that manage access; content describes the array of transportation and payment data and software provided via firms.

The body of work to date on the political economy of social media platforms can be productively compared to the information systems contracting with agencies. Where in social media platforms private governance is located in terms of service and end-user license agreements; in contracting relationships between governments and their partners and vendors, a vendor agreement outlines each parties' rights and obligations. These share a set of resemblances in that "most terms of service include clauses about the platform owner's right to use or sell (meta) data provided by users; few terms of service define the rights of users to access their data" (van Dijck 2013). In the agreements and negotiations I considered, public agencies were cognizant to make ownership and access claims over data—however, the larger question of how agencies' access rights are managed and made real remains. In the conclusion (Chapter 6), I reflect on what lessons learned my findings could offer this broader conversation, and argue that data ownership is a

hollow basis on which to assert and defend users' interests in their data given its failures to do so in my case studies.

Finally, although it is not central to my analysis, political economic explorations of information systems read data as an increasingly valuable commodity. This bears mentioning here as the backdrop for contested claims to data, especially in Chapter 5 which explores behavior from rideshare companies Uber and Lyft. Indeed, accreting granular data is itself a business model given its multiple markets, uses, and re-uses for data (Thatcher, O'Sullivan, and Mahmoudi 2016). This view invites further exploration of the relationship between the material interests at stake in data sharing and access with their social consequences.

> *"[As] Andrejevic concludes, 'People are palpably aware that powerful commercial interests shape the terms of access that extract information from them.' ...These systems have become infrastructural parts of our lives, our cities, and our societies. All without most people having any idea …what the consequences are." (Sadowski 2016, 24; see also Andrejevic 2014).*

Even as my cases focus on the shape of specific data sharing requests, many of these requests are arguably tied to larger, public interest-driven accountability and oversight purposes; it is these goals that are at stake when access to firm data is denied.

3.10 Data Collection

My data collection combines interviews, documents, and meeting notes about data access and sharing across two cases of public agency contracting in the transportation sector, (i) a vendor relationship and (ii) a public-private partnership. Cases were selected based on those where productive conflict could be made more visible as part of a case selection strategy explained in Section 3.2. Altogether, my study consists of 30 semi-structured and open-ended interviews with people selected for their centrality to the cases in this study (Weiss 1995), participant observation of two years of relevant meetings (Aktinson and Hammersley 1998), and document analysis of relevant artifacts (Bowen 2009). In the course of the broader contextual field work that informed this dissertation, I conducted more than 40 formal 60-75 minute interviews with activists, advocates, public employees, and private sector employees who work on government information systems and closely related areas such as privacy, procurement, public records, public-private

partnerships, open data, and open records. Having selected my case studies after years of preliminary groundwork, I held 10 total interviews for 40-120 minutes each across the public employees at the agencies at the center of each case, namely King County Metro, Sound Transit, LA Metro, and ORCA (see Appendix A for example interview protocol). I interviewed one tech activist involved in a public records case regarding code for ORCA card reader devices for 90 minutes, and one researcher involved in both case studies for 4 total hours across 3 interviews. I interviewed a staff person at Vix, the vendor in case study 1, but was unable to use this data because of the employee's preference to speak on background.

My field observations were based in part on my role as a research assistant on a university-based initiative to build a repository for public and private sector transportation data. This experience included my presence and participation in planning policy and strategy for the data repository in numerous meetings on campus, in the City of Seattle, and in representing the Transportation Data Collaborative as a speaker at the 2018 Shared Use Mobility Summit in Chicago, Illinois, which I report on in Chapter 5. My participant observation also included my attendance within local Seattle and Washington State public fora and open public meetings, including numerous events between 2014-2019 including those related to municipal data, open data, public records, civic hacking, data privacy, surveillance, tech activism, civil liberties, technology firms, tech workers, public-private partnerships, mobility, transportation services, technology policy, and urban innovation. These include meetings but are not limited to meetings of Open Seattle (formerly Code for Seattle), the City of Seattle Open Data Camp, the MetroLab Summit, Tech Activism Third Mondays, the Sunshine Committee, annual meetings of the Washington Coalition for Open Government, and the Shared Use Mobility Summit. Over this time period, and particularly between 2016-2018 during my work with the Transportation Data Collaborative, I honed the scope of this research. Once I had selected my case studies in 2018, I collected relevant documents through online search and public records requests including vendor agreements, emails, court documents, and news articles related to each case study. From 2018-2020, I also conducted a series of public records requests in both cases to characterize the process and rationale by which documents are withheld from disclosure in each of my case studies firsthand (see Appendix B for the language of requests).

Not including interactions over the course of the Transportation Data Collaborative effort, formal interviews over the course of this study included (some respondents requested that their job title not be used):

Public employees:

1. Program Administrator, ORCA
2. ORCA Site Administrator, ORCA
3. Senior Enterprise Architect, Sound Transit
4. Employee, Sound Transit
5. Information security employee, Sound Transit
6. Transportation Planner, Los Angeles Metro
7. Innovative Mobility Program Manager, King County Metro
8. Employee, City of Seattle Department of Transportation
9. Employee #2, City of Seattle Department of Transportation
10. Chief Privacy Officer, State of Washington
11. Contracts attorney, State of Washington
12. Attorney, City of Seattle
13. Employee, City of Seattle Information Technology
14. Employee #2, City of Seattle Information Technology
15. Employee #3, City of Seattle Information Technology
16. Employee #4, City of Seattle Information Technology
17. Employee #5, City of Seattle Information Technology
18. Employee #6, City of Seattle Information Technology
19. Employee #7, City of Seattle Information Technology
20. Employee #8, City of Seattle Information Technology
21. Employee #9, City of Seattle Information Technology
22. Employee, City of Seattle Police Department
23. Employee, City of Seattle Office of the Mayor
24. Employee #2, City of Seattle Office of the Mayor
25. Public records attorney, Municipal government
26. Public records attorney #2, Municipal government

27. Public records attorney, private practice

Advocates and activists:

1. President, Washington Coalition for Open Government
2. Member, Washington Coalition for Open Government
3. "The Anonymous Requestor," independent activist
4. Lead organizer, Open Seattle
5. Tyler Traver, the activist who requested ORCA data
6. Member, Seattle Privacy Coalition
7. Member #2, Seattle Privacy Coalition
8. Member #3, Seattle Privacy Coalition
9. Member #4, Seattle Privacy Coalition
10. Member, Center for Open Policing
11. Member #2, Center for Open Policing
12. Member, Safe Utility Meters Alliance Northwest
13. Technology & Liberty Director, American Civil Liberties Union of Washington

Private firm employees:

1. Finance and Operations Manager, Vix
2. Employee, Acyclica
3. Employee, Landis & Gyr
4. Employee, Microsoft
5. Employee, T-Mobile

Researchers and journalists:

1. Director, Washington State Transportation Center
2. Civic technology reporter, News outlet

Events attended over the course of this fieldwork included:

1. Seattle Tech Activism Third Mondays, Seattle, WA 2014-2016
2. City of Seattle Open Data Camp, Seattle, WA 2016
3. MetroLab Workshop on Big Data and Human Services, Seattle, WA 2016

4. Washington Coalition for Open Government Annual Meeting, Olympia, WA 2015
5. Washington Coalition for Open Government Annual Meeting, Seattle WA 2016
6. Open Seattle, Seattle, WA 2015-2016
7. Seattle Department of Transportation Mobility Summit, Seattle, WA 2017
8. Sunshine Committee Meeting, Olympia, WA 2017
9. Shared Use Mobility Summit, Chicago, IL 2018
10. Shared Use Mobility Center, Mobility on Demand Sandbox Summit, Chicago IL 2018
11. ORCA Joint Board meetings, Seattle, WA 2017-2018

Documents received via public records requests included:

- Funding proposals
- Federal grant applications
- Data sharing requirements
- Design of vendor systems
- Patents, user documentation
- Design specs for prospective systems
- Formal policies for data handling
- Public disclosure guidelines
- Privacy policies
- Vendor bids and agreements
- Security audits
- Public records requests submitted myself and found via MuckRock, a website for requesting government records.

3.11 CONCLUSION

In this chapter, I have explained the methodology and rationale behind my choice of ethnographic methods for data collection and analysis. I have also explained how specific traditions in ethnographic methods--an ethnography of infrastructure and conceptual inversion--afford me the ability to include both sociotechnical infrastructures and sociological theories in my analysis. My

interest in "tracing data" in transit is also related a distinct but compatible method known as actor network theory. While these methods emphasize different affordances, they work together to allow me to integrate system design, organizational boundaries, and matters such as legal agreements and policy into my analysis. My research design draws on ethnographic field notes, documents, and interviews to trace data (and its attendant agreements, ownership claims, access requests, and accountability purposes) in three case studies. While my status as a participant observer on one of the cases has enrolled me in some of the same ambiguity about what information I can disclose as the people I study, it has also given me access to the organizations affected and knowledge of what is considered to be at stake in data sharing frictions: privacy harms, transportation equity, environmental concerns, and government's ability to improve service delivery.

Chapter 4. CASE STUDY 1: THE ORCA AGENCIES-VIX VENDOR RELATIONSHIP

This chapter is a case study reporting on fare card transaction data collected by a vendor called Vix Technology ("Vix") on behalf of a partnership of seven transit agencies called ORCA. The purpose of this chapter is to explore the frictions that the agencies encountered in obtaining access to data that contractually belonged to it. While the initial agreement between the partnership of agencies and the vendor specified that the vendor owned the executable code underlying the fare card system, the agencies contractually owned fare card data. Nevertheless, constraints imposed by the original system design ultimately required the agency to pay additional money in order to secure access to this data. I describe how these frictions contributed to the agencies' decision to replace their existing "single provider" contract with a more interoperable "Next Generation" vendor ecosystem. I go on to explore the other actors who wanted access to ORCA data (e.g. institutional customers, law enforcement, transportation planners) and the ways their access to data is negotiated through the Public Records Act and ORCA agency policy. I also describe how one records request for Vix software was denied on the assertion it would violate the vendor's information and trade secret. Taken together, this case study illustrates the multiple ways that ownership and access claims are asserted over data and code (by agencies, by firms, by a vendor, by activists) and how these claims are negotiated under constraints imposed in practice by technological frictions and the law.

4.1 ORCA Regional Fare Card

4.1.1 *History of the ORCA fare card program: its vision and initial implementation*

Transit in the Seattle metropolitan area today is supported a single transit fare card, called the "One Regional Fare for All" Card, or "ORCA" card, which facilitates transit payment across modes of transport and nearby counties. Transit riders can spend $5 on a blue plastic pass card (see Figure 1) at a kiosk or be provided one by their employer, or by social programs that offer discounted passes to seniors and low-income people. The fare cards are loaded with value online by rider's respective employers or by riders themselves at ticket kiosks (see Figure 2). To pay for a transit trip, riders must hold the blue fare card in front of a yellow and black sensor at entrances to trolley and light rail stations (see Figure 3) or a gray sensor at the entrances to buses (see Figure 4).

Holding the card in front of the sensor, a loud beep signals that the sensor has registered the rider's card as a "tap," deducting the cost of the fare from the rider's card and registering it as a trip on the ledger contained on my card. Buses do not have sensors for riders to use on exiting; on light rail trains, sensors at the exits are used in lieu of turnstiles. Any person who forgoes tapping the card on the sensor in the light rail stations runs the risk of being stopped by a Fare Enforcement Officer, who conducts random checks to make sure that riders are paying for trips by tapping into (and out of) stations.

While the single fare card across transit modes is intended to promote a sense of seamlessness in moving across the Puget Sound Region by transit, behind this system lies a complex system for coordinating payments to each separate transit agency participating in this partnership. Buses are managed directly by individual transit agencies, making it easier to allocate payments within and between transferred bus fare rides. The light rail train, however, connects multiple transit jurisdictions together, necessitating more specific information about where riders alight and exit in order to allocate payments. The term "ORCA" then refers not only to the fare card, but also to this complex inter-agency partnership and organizational apparatus that integrates public transportation and co-ordinates payments across different services.

	Timeline:
1991	Vision for regional fare card
1996	Puget Pass program
1999	Original Call for proposals issued after "the region got together and said we need a single method for ticketing that ideally phases out paper tickets…" (interview, System architect, March 28, 2018)
2003	ORCA agencies sign on to vendor agreement with ERG (Contract 229944, see excerpt in Appendix C)
2009	First rollout of ORCA card program
2017	The agencies sign an expansion of the original contract (Amendment 400, see Appendix D) paying vendor $250,000 for the work necessary to provide a copy of their warehouse data to them
2018	Tyler Traver requests Vix software

2019	Vix is granted a permanent injunction withholding software source code from disclosure

Table 5 Timeline of key moments in ORCA Card data sharing

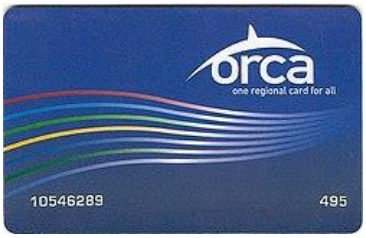

Figure 3. ORCA fare card

Figure 4. Kiosk for buying and adding value to ORCA cards

Figure 5. Sensor for "tapping" ORCA cards on entering and exiting light rail stations

Figure 6. Sensor for "tapping" ORCA cards on entering the bus

4.1.2 *Governing the ORCA inter-agency partnership*

A primary aim of a regional pass program is to promote transit use by integrating payment between different transit agencies. Before the ORCA card, long-distance commuters were especially affected by frictions in transferring rides between distinct agencies. By 1991, transit agencies in the central Puget Sound region began to formalize a shared vision to create an integrated pass for multiple transit agencies. By 1996, the effort resulted in a regional fare policy and a single fare card called Puget Pass, later to be updated and re-named One Regional Card for All.

The ORCA partnership currently spans 10 regional agencies in the Seattle metropolitan area Sound Transit (Link Light Rail, the regional commuter train the Sounder, and express bus service called the ST Express), King County Metro, Pierce Transit, Community Transit, Kitsap Transit, Everett Transit, Washington State Ferries, King County Water Taxi, Kitsap Ferries, and Seattle Streetcar. The partnering agencies in the ORCA card system range in size, in part because the wide coverage area of the ORCA partnership. The largest agencies are headquartered in Seattle; these are Sound Transit, which runs express buses, a commuter rail service, and the light rail; and King County Metro, which runs most Seattle-area buses. The remaining agencies hail from relatively less densely populated areas; for example, Kitsap Transit manages buses for the relatively small population of Kitsap County. The promise of the single card system included a range of benefits, such as improving access to transit and utilization by increasing payment convenience and increasing the interoperability of different transit modes for commuters. However, the partnership would also require a joint governance structure to administer revenues and other assets across participating public agencies.

In order to execute authority across the partnering agencies, ORCA relies on a Joint Board structure, which assembles a group of directors with representatives of each partnering transit agency to meet on a monthly basis and make shared decisions (See Fig 5 for an organizational chart of the ORCA partnership). Many people serving in coordinating roles are employed by the partnership, such as the Regional Program Administrator and Operations Manager. Other people serve in dual roles in which they are primarily employed by one of the partnering agencies (in particular, King County Metro or Sound Transit) and take on duties for the joint board. The partnership is governed by a Joint Board of representatives from each agency. Reporting to the Joint Board, an Operations Manager is responsible for staff overseeing operations, administering programs, and managing business for each agency. Other teams reporting to the Joint Board

include new initiatives like the Next Generation ORCA project, and ORCA-wide security, legal, fiscal, and marketing staff.

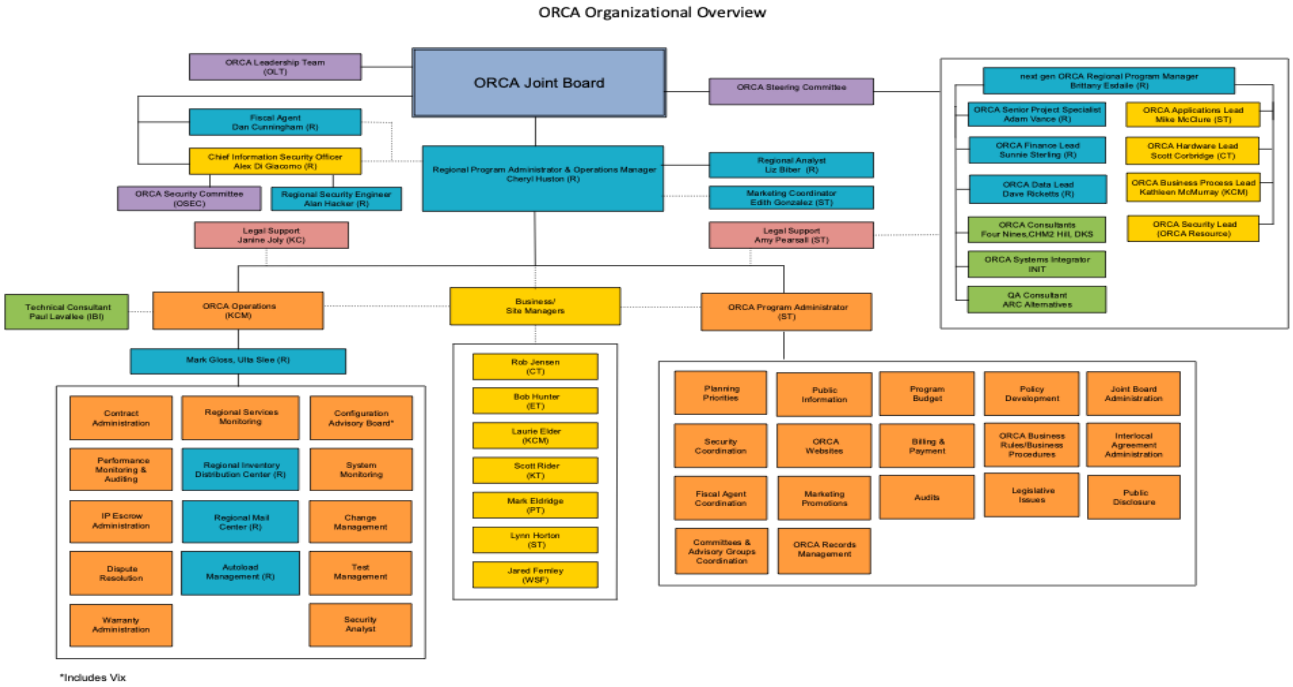

Figure 7. Organizational Chart for the ORCA Partnership obtained via public records request in June 2019

4.2 ORIGINAL CONTRACT WITH VENDOR VIX

The 1999 original call for proposals to create the ORCA card was released at a time before many other metropolitan areas had integrated fare card payment systems. [22] Indeed, few companies were prepared to respond to this initial call. One of the primary selection criteria the agencies used to decide on a contractor was that a firm should be prepared to offer a "turn-key solution," that is, the firm would be able to provide sensor hardware, data collection and warehousing, front-end payment management to customers, and back-end software for agency staff. An ORCA System Architect explained the motivation for a single vendor at the time as driven by "fear, doubt, unwillingness to-- the inability for agencies to do these large systems by themselves and desiring

[22] "In 2002 the region got together and said we need a single method for ticketing that ideally phases out paper tickets. It went to tender and a company ERG now called Vix won, and it was a single vendor run system. That's what they proposed, that's what the competitive process ended up with." (interview, ORCA System Architect, March 28 2018).

for a single throat to choke when things go wrong. Why do we outsource it? Because we needed one person to do it all, instead of rotating" (interview, ORCA Next Generation System Architect, March 28, 2018). This quote points to the close reliance of the agency on the vendor Vix and the original intent in locating responsibility for the system in a single point of contact.

The company that won the bidding process provided an integrated, single-vendor solution called ERG Group. Today known as Vix Technology and headquartered in Perth, Australia the company is incorporated in California to have a U.S.-based subsidiary. In 2018 filings for a court case described later in this chapter, the firm says of itself:

> *"Vix is part of a corporate family that is an industry leader in transport ticketing, implementing and managing automated fare collection, payments, access and passenger information systems for transit agencies cross the world… Representative customers include, in addition to the seven transit agencies who use the ORCA card in the Puget Sound Region, transit agencies in Salt Lake City, Utah; Dallas, Texas; Dublin, Ireland; Rome, Italy; Beijing, China; and Cape Town, South Africa" (August 6, 2018 in Vix Technology (USA) Inc. vs. Puget Sound Regional Transit Authority and Tyler Traver).*

At the time, it had already won bids to provide transit card infrastructure to several of the aforementioned transit agencies, including the Bay Area Rapid Transit in San Francisco.

ORCA signed the agreement with Vix's predecessor ERG in April 2003 and prepared for its system to become operational in 2009. The accepted bid proposed that Vix would provide an "all-in-one" solution including designing, building, and maintaining the ORCA card system, algorithms to direct how fares are distributed among partner agencies, front-end systems to allow card holders to add value to their cards, and ensuring the information security of systems over time. Due to the constraints of technology available for these purposes, the system to be implemented was described to me as relying primarily on data warehousing methods from the early 2000s.

This agreement, formally titled the "Contract for the development, implementation, operation and maintenance of the regional fare coordination system" (Contract number 229944) describes the terms and conditions behind the scope of work the vendor would provide (such as each party's responsibilities, agency's authority, schedule, and intellectual property), the specifications for the services and equipment being procured, and attachments clarifying the price schedule, implementation schedule definitions, and so on. In it, the vendor agrees to provide a wide array of the components necessary to collect and process transit payments, including fare

cards, equipment for scanning cards (referred to as fare transaction processors (portable and standalone), a system for wirelessly synchronizing payment data, a customer service terminal, a data collection system, and a "back office data integration system" for warehousing data. It also agrees to provide a series of services to ensure the system would be successfully implemented, including training and maintenance. Since the contract was initially signed in 2003, the contract has been updated with over 400 change orders (i.e., amendments to the contract).

The vendor agreement specifies that while most aspects of the system are intellectual property belonging to the vendor. Specifically, the contractor IP included the "compiled executable code for each program, module, and element of the software and firmware…the compiled executable code of any third party software incorporated into the contractor IP, security algorithms, user documentation," and other documentation describing the "uses and functions of contractor IP" (Contract number 229944, section 35.4.3, page I-56, see Appendix C).[23] The agreement grants the agencies a nonexclusive, royalty-free license right to use contractor IP for the period of the contact and thereafter (page I-56, section 35.4.3 of contract 229944, see Appendix C). I encountered this assertion of vendor IP directly in this research.[24]

While the vendor asserted that the software and documentation was their intellectual property, importantly, the initial agreement is explicit in that data generated by using the fare cards

[23] This section of the agreement prefigured challenges I faced later in gaining access to information about the system and how it worked, hearing that documents I was interested in were proprietary.

[24] Even as the agencies contractually owned data generated as a result of their contracting with Vix, there were multiple aspects of the system that the vendor asserted to be proprietary, including hardware sensors, software (e.g. for internal reporting software applications), and user manuals. During an interview with a Site Administrator I asked for documentation on the Oracle SAS software used to better follow what she was saying about "universes" and fields---different views available in the system to customer service personnel. She responded: "You would have to make a public disclosure request, and I don't even know if we would be able to give it to you... All of our data is proprietary. All of our design documents... On the Vix site, everything is proprietary, so I would think that ... Well, I don't know. I mean it could be that... You'd have to get a nondisclosure agreement from Vix." Some of the ambiguity about the reach of Vix's proprietary rights over ORCA work products is attributable to the way the Public Records Act works. Whether a particular record is disclosed is a function of a court determination, which provides third party notice to parties who could file an injunction on the grounds that requested information could harm their competitive advantage.

belongs to the agencies. More specifically, it states that use data[25] and any use data reports[26] generated in the course of business would be the "sole and exclusive property of the Agencies" (Page I-56, Section 3.I-35.7.1 of Contract 229944, see Appendix C).[27] The vendor was to be granted a non-exclusive, royalty-free license to use this data for the purpose of executing their work. Moreover, the contractor agreed to provide this data to the agencies on request, or if sooner, the expiration or termination of the contract. Nevertheless, a respondent shared, "Though [we had] all the usually contractual protections for things like, "The region *owns* the data, *that doesn't mean that we have access to the data*, as has been found out recently..." (emphasis added, interview with ORCA Next Generation System Architect, March 28 2018).

While the agreement was clear that the agency would own the data and indicated the vendor is required to provide it on request, the system in practice presented technological constraints that prevented the frictionless export suggested by the language in the original contract. By 2017, the agencies ultimately signed an expanded scope of work to export its data from the contractor's data repository. This amended scope of work cost the agency for $285,182.[28] My initial foray into this

[25] Officially, use data is defined as **"**Data and other information of all types, whether or not copyrighted, that is created, generated, collected, discovered, or otherwise obtained by the Contractor, or its subcontractors, in the course of performing the Work, and that relates to the use of RFCS Cards, the RFCS Application on non-Agency issued cards or media, or the operation of the RFC System. Examples include, but are not limited to: data related to customers of smart cards, data related to the sale of smart cards, data related to the addition or subtraction of value to or from smart cards, data related to the revocation of smart cards, data related to the linking of a particular smart card to an individual person, data related to the use of smart cards, and any data generated from the foregoing" ("Definitions," Contract 229944).

[26] The Contact defines use data reports as "Any summaries, manipulations, extractions, representations or other reports of Use Data created by the Contractor, its Subcontractors, the Agencies or any other entity authorized by the Agencies." (Contract 229944 Exhibit 01- Definitions).

[27] Specifically, the agreement reads "All Use Data, Use Data Reports and all rights All Use Data, Use Data Reports and all rights thereto, shall be the sole and exclusive property of the Agencies. Contractor hereby irrevocably assigns exclusively to the Agencies and their successors and assigns any and all right, title and interest in the Use Data, including all copyrights, trade secret rights, and other proprietary rights, title, and interest thereto. For clarification, any Use Data Reports that are created by the Agencies shall be the property of the Agencies. … The Agencies grant the Contractor a non-exclusive, royalty-free license to use the Use Data for purposes relating to the performance by the Contractor of any of its obligations under the Contract. The Contractor shall at all times maintain the privacy and confidentiality of the Use Data and any Use Data Reports created by the Contractor or its Subcontractors. All Use Data and Use Data Reports shall be at all times segregated and kept confidential by the Contractor, and may not be used or disclosed, in whole or in part, in any manner except as expressly authorized by this Contract or with the written consent of the Contract Administrator. All Use Data and Use Data Reports shall be returned to the Agencies upon request or, if earlier, upon expiration or termination of this Contract. To the extent that such materials consist of copies of information also in the Agencies' possession, Contractor may alternately certify in writing to the Agencies that the materials have been destroyed.

[28] The amendment to the initial agreement that accounted for the extra costs to extract the data is summarized as "[Agreement for] the Contractor to facilitate the migration of ORCA usage data to an internal ORCA Agency data repository to help validate and drive the next generation ORCA reporting requirements" (Page 57, Exhibit 9 of Contract 229944, Amendment 400, see Appendix D).

case study was based on the understanding that the agency's data access was barred when the vendor asserted that ORCA data to be proprietary. Instead, a closer look demonstrated that the frictions to data access were *not* contestations over who owns the data, but rather constraints imposed by system design. To this point, the ORCA Regional Program Administrator explained:

> *"So, from the data perspective, I don't know what Vix or [its Operations Manager] would have told you about that particular program but it is an archaic platform. The foundation of ORCA is built in an environment that you wouldn't find in today's world, but we're transit, and we can't just say, 'Oh, let's go spend $100 million and replace it.' You have to ride it out. And it is working for customers, it's just not as efficient to pull data and obtain data in useful formats, so when we build the new one, believe me all these lessons learned will pay off" (interview, ORCA Regional Program Administrator July 11 2018).*

I describe the choices that the agencies made in revising ORCA later on.

This case points to the ways in which data ownership is not dispositive of access and control over data. One respondent pointed to ownership as more performative of data governance than acted upon: "A lot of the decision-making has [to date] come from a point of minimum rights protection: 'We own the data.' 'Oh good.' The end," (interview, ORCA System Architect, March 28 2018). Another described Vix as a source of delay: "We are beginning to work on the contractual arrangements with Vix to get access to our own data; technically they can't withhold from us but they can put a lot of red tape around it" (interview, March 29, 2017). In the following section, I explain in more detail how system designed presented constraints to how data could be warehoused and exported. I go on to describe the agencies' emergent interest in "having the data and trying to do something with it," and the large-scale project they undertook to do so (Ibid.).

4.3 ORCA System Design

The current system turns on a key constraint; namely, that the system of record is on the transit card itself. ORCA was designed at a time when real-time communication was not robust enough to support wireless payment infrastructure. As a result, the system relies on the fare card to behave as an authoritative ledger; rather than a database tracking the total value assigned to a card ID number, the card itself stores information about the total value and most recent transactions. "It's truly awkward, there are a lot of workarounds. The technology back in 2002 or 2003 when this

was first designed, card-based systems [i.e. systems in which the transaction is stored on the card] were really all you could get. The communications backbone wasn't there to support real time communication" (ORCA Next Generation System Architect, March 28, 2018).

The card stores information about:

- the user's trip history,
- current balance,
- a list of transactions,
- reference IDs for locations where card was tapped, and
- the last 20 transactions the rider made with the card.

When the rider taps a card on the sensor at the front of the bus known as a "validator," the system performs a fare calculation based on the fare type of the rider (e.g. standard, student, senior citizen), the mode of transportation and agency it belongs to, the time of day, and which route the rider is on. Each tap synchronizes the information on the validator with the information on the card.[29] While the card is considered the authoritative record of a rider's transaction, the validator also stores a log of whose balances have changed; at the end of the day it uploads this information over wi-fi to a database controlled by Vix called the "Back Office."[30]

Importantly, the agency's challenges in accessing its historical data were attributable in part to this aspect of how the data warehouse was designed. The Back Office is a data management system from which the agency is only provided the most recent 25 months of data in the form of a reporting database. While the agency can query the reporting database, agency employees have no access to the Back Office controlled by Vix, which contains both recent and historical data (including data not contained on the reporting system, such as the equipment logs and data older than 25 months). The Back Office uses an approach called Online Transaction Processing (OLTP):

> *"OLTP systems are notoriously hard to query. It's just a fact of compromise of design. It's designed to write very quickly, so it's set up in a way where you*

[29] The system architect recounted, "when the operator starts up the bus in the morning, the validator [hardware on which users tap their transit cards] gathers all the information it needs [such as] who has an adjustment to their balance?... That gets updated to this reader, and in fact every single reader in the entire fleet, all 2000 of them get that "Meg put 10 dollars on her card" the next day." (ORCA Next Generation System Architect).

[30] When the bus comes back to the bus yard at the end of the day, it is connected to wi-fi for a 10-minute window of time in which it uploads its transaction data and receives more updates before powering down; "The yards have wi-fi dead spots, or operators can drive out of range before the process is finished, so it can take a couple of days to get this information onto that reader."

> *can just dump information and walk away. If you're trying to make it user accessible on the other end, you've got to… hit a lot of databases. It's very taxing on this sort of a system. So, most reporting systems have a different sort of schema model that's designed to be slow to put the information in, but fast to read it out" (interview, ORCA System Architect, March 28, 2018).*

The agencies are provided information from Vix via a reporting system, which is connected to the Back Office and is designed using a similar architecture. To improve performance, the reporting system draws only from the last 90 days for reports generated-- up to maximum time span of 25 months. "The Back Office captures the data, and then it goes to the report server. The report server is a window into the database basically" (interview, ORCA Site Administrator, interview June 25, 2018). Between 2009 and 2016, the agencies and the vendor partnered to make improvements to the system that improved its usability.

This system design provided agencies with data access for routine purposes. Indeed, roles ORCA and partnering agency personnel, like the ORCA Regional Program Administrator and the ORCA Site Administrator, described that customer service data needed on a day-to-day basis was readily available by Oracle and accessible via the reporting system, which provided data from the previous 25 months.[31] In fact, one of the people I interviewed reported that she had no problems with accessing data from Vix, to the extent that I thought that I had chosen my case study based on a misunderstanding and I would have to look for a different example of agency-vendor data sharing to explore in my dissertation. As I interviewed additional actors, I came to understand the data access needs for law enforcement and especially transportation planning required more costly requests in the form of change order to Vix, which took a long time for agencies to receive due to limited staff available to fulfill these requests at the vendor.

One administrator employed by the agencies explained:

> *"[We have] access to the bulk of information...anything to do with cards or transactions... There's ridership data. There's financial data. There's maintenance data, when did a device last connect, data on cardholders... There's just not a lot of data that we [regularly use and don't have]--except for the archived data, data that's older than two years-- …that's the kind of data*

[31] In addition to limiting the timespan of information available, the reporting system also constrains the kind of information that can be extracted from the system. As the Site Administrator explains, ""I mean it's not like ... We have what's called universes, and within those universes there are fields, and you can combine fields. In a singular report, we can do ... But if it's something that doesn't reside ... If it's data that we need that doesn't fall within any of those universes or fields, then we can't get it." (interview, ORCA Site Administrator June 25, 2018).

that we have to ask Vix directly for" (interview, ORCA Site Administrator June 25 2018). "

To access data collected more than 25 months ago (for instance, in the event of a law enforcement subpoena, search warrant, or customer service matter for data older than two years), the agencies relied on work requests to Vix:[32]

> *"We just submit a form, if it's archived data. Vix isn't, oftentimes, Johnny on the spot, but we get the data, and it's sometimes not as timely as we'd like it. And if it's for a search warrant or a subpoena, then they comply within the amount of time, and it's generally a few days. So, if it's got a fast turnaround, they can comply. It's not as easy for them to do that, but they can." (interview, ORCA Site Administrator, June 25, 2018).*

Due to the OLTP architecture of the system, querying the Back Office system is time and resource intensive; personnel at Vix who **were** responsible for providing data access requests had competing demands on their time. One Sound Transit employee explained, without wanting his job title to be used, "Because the model is so convoluted and cumbersome today, that's why data is not used to the full extent. One of the problems has been that it is already convoluted enough for the basic processes we need, why even bother [doing data analysis on] something else" (interview January 31 207); he went on to add that specific terms Vix required for third parties to be PCI compliant to be interoperable with Vix systems also reduced third parties' interest in interfacing with the existing data system.

While data the vendor provided was enough for routine needs, over time the agencies' expectations for what they could be doing with data expanded. "It's at a pivotal change in perception right now…. *Owning* the data used to be important. Now *having* the data and trying to do something with it is seen as important. They still don't know what, they still don't know how" (emphasis added, ORCA Next Generation System Architect, March 28, 2018). Respondents' perspectives on data access from the vendor also varied with respect to their role. Whereas personnel responsible for customer service felt that data was readily available, those interested in using ORCA data for long term planning felt that the data available was ill-suited to their needs; it was difficult to re-use ORCA data for planning purposes.[33]

[32] Law enforcement also makes access requests for ORCA data. Under state law, the ORCA agencies require a court order to disclose individual travel records. Law enforcement requests for data are handled by the site administrator, who says they are received about once a month, or occasionally 3-4 at a time.

[33] I came to understand that "taps" were not already used in a robust way for transit planning. Over two summers in 2017 and 2018, a team of students at the UW Data Science for Social Good Program worked to clean transit card

From a transportation planning perspective, intensive and exacting data cleaning and processing would be necessary to make the data useful for long-term analyses. After an extended period seeking access to historical ORCA data, Mark Hallenbeck, Director of the Washington State Transportation Center (TRAC) at UW, secured access to an initial dataset that included 21 million data points from 2 months of "taps" on bus sensors and route schedule information. Driven by an interest to make ORCA data fully usable for longitudinal analysis and planning purposes, Hallenbeck led two teams of student data science fellows at the eScience Institute Data Science for Social Good program (UW Data Science for Social Good Program 2017). These research teams found that ORCA data required many months of cleaning, given that the system was not designed to support such data analysis. For example, different bus route datasets used different map projection systems, such that initial maps of the data placed bus routes "in the middle of Lake Washington." The datasets also included no data dictionary or other metadata, nor documentation of how they were generated by vendor hardware systems. Vix engineers did not respond to the research teams' requests for information; Hallenbeck's team sought to understand many attributes by conferring with system architects across at least 3 agencies. Over the course of many months, the team cleaned and prepared the data to be more useful for planning-- for example, to gain insight into the most popular bus routes, patterns in transits, and an analysis of the effectiveness of programs to increase commuter uses, as Hallenbeck et al. report on "Use of Electronic Fare Transaction Data for Corridor Planning," which describes the research team's experience using the data in depth (Hallenbeck et al. 2017).

---

data to prepare it for analysis. My surprise was driven in part by the imaginary of frictionless surveillance that had originally motivated me to enter this field. I had a vivid memory of the documentary by journalist Laura Poitras, Citizenfour (Poitras 2014), a film which excerpts the following remarks from former UW student and co-founder of Seattle Privacy Coalition Jacob Applebaum saying that the government could create detailed location trails of citizens using their transit card data—a claim that stayed with me over many years:

> "Anybody see how the subway link your metro card to your debit card, right? And, like, auto refill. This is a concept which is key to everything we'll talk about today. And it's called linkability. Take one piece of data and link it to another piece of data. So for example, if you have your metro card and you have your debit card, you have those things and you can draw a line between them right? So that's, like, not a scary thing. Except your bank card is tied to everything else that you do during the day. So now they know where you're going, when you make purchases. So when they decide to target you, they can actually recreate your exact steps. With a metro card and with a credit card alone. Like literally where you go and what you buy and potentially by linking that data with other people on similar travel plans, they can figure out who you talk to and who you met with. When you then take cellphone data, which logs your location and you link up purchasing data, metro card data and your debit card. You start to get what you could call metadata, an aggregate over a person's life" (Applebaum 2012).

I was surprised, then, to learn that at least for Seattle's transit card, these location traces were only being prepared for data analysis many years after system adoption, and even then, only for transit planning.

Researchers' experiences trying to make Vix data usable prefigured the ORCA agencies' own experiences in trying to warehouse the data provided to them by Vix. As the Regional Program Administrator described it:

> *"The challenge we have is taking raw data from Vix and now we're putting it into a data warehouse but it's like wow, it does not-- that apple is not fitting into our apple crate. We cannot—so—Vix has to provide [us with] more definition—and definition is not the right word, it's just the way these files are built, again, frankly built in 2006-2007, so not in current standards. They can extrapolate it but they have to do a lot of manipulations. So, the taps are all there …but it's not an easy thing for us to take their raw data and do anything with it without a lot of work" (interview, ORCA Regional Program Administrator, July 11 2018).*

Experiences like these point toward the laborious process of making meaningful use of data once researchers and agencies have secured access to it, and the ways in which data access alone is not sufficient for agencies to make data useful to public needs.

4.4 NEXT GENERATION ORCA

4.4.1 *Factors driving the move to Next Generation*

The constraints imposed by using the fare card as the system of record created the conditions that ultimately led the agency in 2016 to begin a large-scale effort to design a "Next Generation" system to improve data access, portability, and interoperability challenges encountered in the existing data system. Next Generation refers to the suite of initiatives that included (i) a new competitive tender process to overhaul the fare card payment system, (ii) planning to promote modularity and interoperability between different components of the Next Generation system, so that the agencies could more easily upgrade some aspects of its system and avoid vendor lock-in, and (iii) a data warehousing initiative that would ingest historical data from the operation of the ORCA system to that point.[34] While much of the existing ORCA system worked reliably and as intended, two "pain points" in legacy ORCA system design drove the agencies to pursue a "Next Generation" system.[35]

[34] As a result of the difficulty in securing data access to date, Sound Transit will be overhauling its enterprise data warehouse with a 'modular' approach that does not depend solely on one vendor; to date, "All the data [has] live[d] on the SCADA side, it's a walled garden" (January 31 2017).

[35] Interviewees described how the Next Generation system would be executed carefully so as not to undermine what was working well about ORCA. "People put money in, they tap cards, that works. And unlike some other cities that

First, riders' reported frustration in reloading value onto cards online; the process of synchronizing online transactions with the fare cards could sometimes take up to 72 hours.[36] As a local transportation planning expert familiar with the system described it in March 2017, "ORCA #1 is brain dead; they're trying to keep it on life support. They're pumping blood to its heart but the brain is gone. Next Gen ORCA, they are still 4 years out. Some parts will come together earlier than that—but [the agencies] have a scope creep problem; because they are making it modular, they need to have a tech spec built with all the ins and outs [of each component]" (interview March 16 2017). This quote reveals that even as they attempted to address the risks of relying on a single vendor, other frictions are presented by the process of making multiple vendors interoperable through detailed specifications. On background, employees at ORCA agencies acknowledged the ORCA program's stability and success while also expressing frustrations with their experiences working with Vix.

Due to the sensitivity and long-term frictions that led to contracting for Next Generation ORCA, the interviews I undertook with both ORCA agencies and its vendor Vix in 2018 were delicate. While Vix was still running the existing ORCA system, the agencies had a then-active call for proposals looking for a vendor to replace it—which Vix had submitted a bid to alongside its competitors. Although I had been granted an interview with the Operations Manager at the Vix office in Seattle, during our interview he was willing to talk but not to sign the informed consent document without having it reviewed by Vix's legal counsel. He had also offered to put me in touch with the Vix system architect responsible for the ORCA system pending the outcome of the legal counsel's review of my informed consent document. On the weeks after our interview, the respondent declined to reply to emails or consent, such that I was not able to use the contents of our interview.

---

have tried these sorts of things, people like ORCA, so, this next generation of ORCA program [goal] is to not destroy any goodwill we have in the community, but also give it all the things that are missing today and solve some of the key frustrations people have with the system and give it a makeover" (ORCA Next Generation System Architect, March 28 2018).

[36] The constraints described above mean that adding value to fare cards online (as opposed to an in-person kiosk, where the rider can scan the card directly) requires a set of processes for reconciling the added value with the ledger on the card; there could be a 2-3 day delay while information about payments added to fare cards is transmitted to the sensors onboard buses; adding funds is not complete until the card is scanned and the ledger updated on the card itself.

4.4.2 *ORCA agencies warehousing data in the cloud*

A key part of the Next Generation effort was the agencies' interest in warehousing their own data in the cloud to make it more accessible to longitudinal analysis. This data warehousing effort led to the aforementioned Amendment 400 to the original contract with Vix, which covered the cost to export historical data from the vendor system. This effort encountered the same frictions that had previously encumbered agency access and use of data; as one ORCA employee put it, "I deal a lot with Vix, I am trying to get data out of them right now as part of the Next Generation program. It's a long hard slog" (ORCA Next Generation System Architect, March 28, 2018). The Next Generation system was intended to make the data more usable: "The ideal solution is that the data warehouse now being selected gives us these options and avenues and ways to consume the data [long-term]" (interview, January 31, 2017).

Early on, before the agencies had a solution in place for its data warehousing effort, our team at UW was discussing whether our incipient data repository effort could assist with their interest in protecting rider privacy in the granular transportation data. In a meeting, we shared our research on privacy oversight and data governance in the context of a large public agency, and offered to continue the conversation on what privacy or security plan each of the partnering agencies at ORCA would need for their interest in stewardship over their historical rider data. As part of this series of conversations, we asked if the UW Transportation Data Collaborative could take on any existing ORCA data assets in order to begin piloting our privacy preservation techniques—as well as explore the possibility of combining ORCA data with other local transportation toward our vision of cleaning and aggregating data across multiple transportation modes to be more useful for transportation planning. In 2017, we presented our repository as a means to help them prepare for Next Generation, or as a pilot to help them identify their needs. By late 2018, public transportation hardware and software service provider INIT won a $94 million competitive contract optimize interoperability in the Next Generation ORCA system.

4.5 Other Parties' Data Access Claims

Beyond the relationship between the ORCA agencies and their vendor, myriad other actors and individuals assert claims over ORCA data. In this section, I explore various stakeholders' data access requests, the way that data ownership figures into these requests, and how they are handled

under the Public Records Act. Recall that ORCA is a payment coordination system; many bids to access ORCA data I describe in this section are tied to payments and budget oversight. The following examples of data requests by interested parties trouble the discrete definitional boundaries between agency and vendor intellectual property as laid out in and adjudicated by the vendor agreement.

4.5.1 *ORCA data access by institutional customers*

Until now, this chapter has focused on individuals who buy and use ORCA cards, but ORCA customers also purchase cards at an institutional scale. Institutional customers include local universities like the University of Washington, or companies like Amazon and Safeway. How much visibility employers are allowed to have into individual use has been a major site of contestation since the inception of the program. Because they pay for ORCA cards, and institutional ORCA card accounts are billed based on usage, employers historically argued for the right to the data generated by these transactions. Employers' primary interest in granular data access was to assess their transit card expenditures for cost savings opportunities:

> *"They wanted to know well, where are our people, which routes are they coming from? When do they use them? How many are actually using them? So, we know that if we are paying for 400 cards, if 3 people are using them, it would probably be in our financial best interest to not offer them anymore. If 50 people are using them heavily and 10 people are using them kind of a bit, that might be a break-even point, you know, they've got their finance people who will work that out" (ORCA Next Generation interview March 28, 2018).*

This quote also speaks to employers' interest in not just in whether to invest in transit cards, but in more specific questions about their employees' transportation patterns geographically and over time. The respondent went on to explain how major Seattle area employers (like Microsoft and Amazon) have decided to provide private transportation shuttles to employees as part of their transportation spending; for decisions like these, granular data access about ORCA card usage could inform the cost effectiveness, geographic distribution, and timing of private shuttle services.

Another major reason that employers desired granular access to ORCA card usage data was to assess whether employees were misusing transit cards. In this vein, institutional customers are interested in whether employees are solely responsible for card use expenditures, or whether share cards with friends and family. Strictly speaking, many employers also intend that their cards

be used only for work commutes, as opposed to leisure time. Finally, a final set of data access requests stem from employers monitoring employee conduct. A Sound Transit employee described these requests as "less well intentioned, like [employers] who want to see if Fred who said he was sick on Friday… [but instead] went somewhere. And we have had those questions specifically asked because there are discipline cases where the employer feels as though they need to support disciplinary action, and there's fraud as well" (interview System Architect, March 28, 2018). It was this more individualized monitoring that most concerned community advocates.

From the moment the ORCA cards program was launching in 2009-2010, the Washington and local privacy advocates voiced concerns that the usage records created by the ORCA card system could be privacy invasive if employers were allowed to access the use logs of specific employees. Among the scenarios they warned about was the concern that data "potentially could end up in divorce and custody cases and other legal proceedings [or for] card holders to be tracked as they use public transit to go to work, church, shop or participate in political rallies" (ACLU Washington 2010). Heightening these concerns, the original data sharing policy with institutional customers was to share the same personal data that individuals could request and receive about themselves:

> *"[Individual users already got] a very detailed "I got on this bus at this time going this direction" location history…. [in the beginning,] institutional accounts started getting access to [the same thing]… And the ACLU said no, that's not cool, you can't let the employer know where people are. There's no ability to get informed consent about granting that; you can't guarantee that any consent is informed and not coerced in any way" (interview March 28, 2018).*

ACLU WA asserted that transit data to be considered personal information, that is, belonging solely to individual riders. Specifically, they argued that employees receiving a transit card have an asymmetrical power relationship with employers and are not able to meaningfully opt into or out of its use. Advocates pushed for employers and schools subsidizing ORCA cards for employees and students to be prevented from accessing how individual fare card recipients use data.

A contemporary, 2007 version of the WA Public Records Act describes the full set of actors and purposes for which detailed data was to be provided.[37] This list is found in the part of the statute exempting individuals' transportation data and financial records:

> *"The following information is exempt from disclosure:…the personally identifying information of persons who acquire and use transit passes and other fare payment media including, but not limited to, stored value smart cards and magnetic strip cards, except that an agency may disclose this information to a person, employer, educational institution, or other entity that is responsible, in whole or in part, for payment of the cost of acquiring or using a transit pass or other fare payment media, or to the news media when reporting on public transportation or public safety. This information may also be disclosed at the agency's discretion to governmental agencies or groups concerned with public transportation or public safety" (RCW 42.56.330(5), emphasis added).*

This language is very clear about allowing employers to obtain personal data, by virtue of the fact that they are *paying* for transit cards.

A 2010 revision to these lines of the Public Records Act substantially amended the amount of data and circumstances under which both employers and law enforcement can obtain detailed data. The revised chapter (RCW 42.56.330) prohibits individual-level transit card information be disclosed to employers, except for the purpose of preventing fraud:

> *"The following information relating to public utilities and transportation is exempt from disclosure under this chapter:...The personally identifying information of persons who acquire and use transit passes or other fare payment media including, but not limited to, stored value smart cards and magnetic strip cards, except that an agency may disclose personally identifying information to a person, employer, educational institution, or other entity that is responsible, in whole or in part, for payment of the cost of acquiring or using a transit pass or other fare payment media for the purpose of preventing fraud. (RCW 42.56.330(5), emphasis added).*

The statute further specifies that information can be shared only when it does not contain any personally identifying information. In cases where a pending investigation requires access to identifiable data, the law provides it can be released to law enforcement agencies when accompanied by a court order. Employers and law enforcement would still be able to request access

[37] Note that much of the original language of RCW 42.56.330(5) exempting transit passes and fare payment media pre-dated the ORCA card by several years; interviewees I spoke with were not familiar with the earliest history of this part of the statute.

to individual transit records under exceptional circumstances: "It's a case by case… [they need to] pull up Fred's history and provide a case. Or there's a crime involved, and police come in and subpoena data" (interview March 28 2018).

This pushback resulted in changes not only to the Public Records Act, but to ORCA agencies' policy whereby the agencies would provide only aggregate use statistics to employers.

> *"So [institutional accounts] just don't get access to that data. What you can do instead is find out how many of your cards were used; [but] there's a constant pressure [on ORCA agencies] from some employers who really want to fine tune their costs. And some of them have good intentions… "But we can't give them that detailed information." (ORCA Next Generation System Architect, March 28, 2018).*

Because ORCA does allow employers to request individual-level records in the event of fraud, the ORCA Regional Program Administrator described how employers seeking robust data access turned around and began "requesting fraud for all [their] cards," to which she responded, "That's highly unlikely… so we've gone back and forth with a number of businesses. It's like 'No.' 'No.' 'No'" (interview, ORCA Regional Program Administrator, July 11, 2018).

Multiple employees within Sound Transit also described the legacy to me of this advocacy and organizing as the ORCA agencies ultimately interpreting the language of the statute very narrowly as to what constituted permissible data sharing with employers. As one Sound Transit employee explained:

> *"Although there is an exemption in the [Public Records Act] for [card-based transit data] it has not been litigated. Our legal has always had a conservative reading of it… The original intent of the exemption was cards as in a credit cards, it's not written to MEAN the ORCA card itself, but it could be interpreted that way. Until there is legal precedent, your guess is as good as mine… That's why our de facto practice has been to take a conservative approach to protect the privacy of the data" (interview January 31 2017).*

This respondent speaks to the seams between the language in the statute and agency policy, and the uncertainty that agencies must operate under when particular data types are not clearly addressed in the statute.

Nevertheless, continuing pushback from institutional customers was leading in 2018 to a re-evaluation of ORCA policies. Today, institutional customers' requests for data are escalated to the ORCA Regional Program Administrator, who described the agency's policy as:

> *"We do not go above Washington State law… We are in the process of assessing or reassessing the policy and [state laws] that may be related to what the businesses can obtain and that is currently… not for public consumption yet… we are having the agencies' attorneys review and reassess what can be provided. Our ultimate goal always is to protect the individual card holder and their data" (interview, ORCA Regional Program Administrator July 11, 2018).*

In this sense, the language of the Washington Public Records Act itself (and the agencies' conservative interpretation thereof through policy) mediated institutional customers' access to data.

4.6 Contesting Intellectual Property Claims with Public Records Requests

Just as the public records act played a role in negotiating institutional actors' access to ORCA data, it also mediated individual access to records requested. I spoke to a local technology activist, pseudonymized as Tyler Traver, who had left local giants like Amazon and Microsoft because of his desire to work in the public interest.[38] On May 5, 2018, Traver had requested the software used on fare-enforcement officers' scanner (specifically "Psion handheld devices" see Fig. 6); officers tap each rider's card on the device and it validate whether the fare had been paid. Before I met with Traver, I had misunderstood his desire to inspect the software as a reaction to being caught and ticketed for riding without paying the fare.[39] I was surprised when Traver told me that he had never received a ticket, nor warning. Rather, his public records request for fare enforcement software was the latest in a series of requests he had made over time that pushed at the original intent of the Act, driven by a hacker ethic of experimenting, breaking, and repurposing:

> *"I saw the Public Records Act as a system of rules, and I like taking systems of rules and making them do unexpected things. I don't think they had software in mind when they were writing the Public Records Act. I don't think they had this type of audio file in mind, so really it was just… originally, I just wanted to do everything I could do get a novel artifact out of it" (interview March 19, 2019).*

[38] Working part-time as an education sysadmin, he said his job was "Not super interesting from a technical perspective, but it's something that I feel a lot better doing than building systems that end up getting used by advertisers" (interview March 18 2019). Previous to my work with the City of Seattle in 2015, I had spent more than a year working with local tech activists and open source advocates, who shared similar personal stories of leaving the large technology firms in the area.

[39] Given the absence of a turnstile like that used in other urban subway and light rail systems, Seattle's light rail currently uses fare enforcement officers who inspect transit riders' fare cards with hardware validators to monitor whether the rider recently "tapped" the card onto the payment sensor.

This interest in requesting novel record types started with the source code for ORCAcard.com, to which the agency responded that they had no responsive records: "I went into that trying to get any software out. It wasn't fruitful" (interview March 19, 2019).[40] Later, Traver moved to request the audio files that play on light rail trains at stations and stops (J. P. Brown 2017). At first, he told me that he was not personally invested in the language and limits of the Public Records Act itself, but instead what he could do with it. That perspective changed over time, especially when his exploratory request garnered an unexpected series of reactions from the agency and firm.

Figure 8 Handheld Psion scanner devices used by Fare Enforcement Officers

Traver had never received a fare enforcement violation but had been subject to multiple checks which occur in Seattle on an ad-hoc basis. He shared, "What got me requesting things about fare enforcement was, whenever I'm on a train and I get fare inspected, my adrenaline—it's a law enforcement obviously, or someone who appears to be one on the train, and that's a situation that I'm uncomfortable being in" (interview March 19 2019). Acting on this discomfort, Traver decided to file a public records request for fare enforcement policies and protocols. In response, the ORCA agencies provided everything except for records withheld on the grounds that they were being implemented in the software called "ORCA Inspection" and "Officer Android" provided by Vix, and that their disclosure would conflict with the vendors' trade secret. Uniformed officers carried card validator devices to scan a rider's ORCA card in order to detect whether the card had been recently scanned to pay for the trip.

While he had not originally planned to submit a records request for the fare enforcement software, Traver followed up by re-asserting his request—this time for a copy of the code for both ORCA Inspection and Officer Android. He recalled:

[40] This case is an excellent example of the agency responding to a request for data handled by a third party in a way that originally interested Jan Whittington and I in 2015—namely, that because the agency did not directly administer a set of records, many public disclosure requests were denied. Depending on the facts of the case, we learned that this clean distinction is not so simply drawn in practice, and, given the increasing amount of public information administered by third-party vendors, could have high stakes for public access and oversight.

> *"This was an opportunity to get software through the Public Records Act which is something that I [already had] an interest in doing…[since] my background is in software, I think the increasing role that software plays in everything is—it's would be interesting to get software just to establish that software is a public record, or can be a public record" (interview March 19 2019).*

Traver explained that he was not only interested in seeing how the law applied to novel types of records not anticipated in the 1972 Public Records Act, but also to understanding software as an extension of (and operationalization of) the law when it is used in the conduct of government business. Later in the interview he drew a connection between his interest in hacking on law and hacking with software:

> *"Law is a code. Software is a code, right? Law tells you what gets done under certain circumstances. Software is the exact same thing. I think that's becoming more and more relevant as software gets more integrated into our government… software [should be] subject to public inspection" (interview March 19, 2019).*

Traver was interested in the public transparency and accountability dimensions of public access to vendor software but faced a vendor with a strong commitment to protecting its intellectual property. I interviewed the ORCA Site Administrator when the case was still in its early stages; she asserted, "Vix isn't keen on sharing information...I would strongly doubt that Vix would be inclined to give them a copy of that because, like I say, they believe all of their software and hardware is all proprietary" (interview, ORCA Site Administrator June 25 2018).

The language of Traver's request was limited to a version of the software that was contained on the fare scanner devices, called the binary code. Traver explained to me the distinction between binary code and source code as a translation from the human readable to the machine-readable version which occurs when source code is "compiled," or translated, to create a binary version of software, which runs on hardware. Fare enforcement officers used Vix scanner devices running the binary version of code but did not have direct access to the source code the vendors used to produce it. As a result, Traver was careful to assert his request for the binary code; recall that his previous request for ORCAcard.com had been denied based on the fact that the ORCA agencies did not directly control the requested record. The distinction drawn here echoes our previous work with the City of Seattle in 2015, where we saw some public employees officers draw a distinction between data that the public agency directly had or managed, and that administered by vendors—asserting only the former to be disclosable.

Traver was interested in gaining access to the source code but believed only the binary code on the officers' scanning devices to be available. In our interview, he shared his hope to leverage the binary code to pursue access to the source code indirectly—by reverse engineering the parser. Traver speculated that the vendor's software could have been based on a well-known app called FareBot, an Android app which allows individuals to scan transit cards with their phones in order to display a card's transaction history.[41] FareBot was released as open source software under a GNU Public License; the terms of this license require derivative software to also release their source code as open source. In this way, inspecting the binary code was one tactic Traver planned to pursue to gain access to the software's source code, saying:

> *"The GNU Public License says that if you distribute a binary version of an application to somebody, you also have to give them the option of requesting the source code from you. The idea behind this is similar to a lot of open source licenses, to spread the idea of … software being available for inspection… If we can get a copy if the binary of the app and it does turn out to be based on FareBot, that would give FareBot authors a hook to request the source code" (interview, March 19, 2019).*

Traver described this strategy as not only tied to his larger commitment to free software, but to his interest in hacking the information on the ORCA card to see what attributes it stored, and contributing anything useful back to the FareBot app "to make Farebot more future-proof" (interview March 19, 2019).

The ORCA's agencies received Traver's request and wrote to Vix that "it is not readily apparent [to Sound Transit] whether an exemption exists under the Washington Public Records Act to prevent Sound Transit from producing" the software (August 6, 2018 in Vix Technology (USA) Inc. vs. Puget Sound Regional Transit Authority and Tyler Traver (pseudonymized)). After the ORCA agencies provided Vix with third-party notice, the company filed a court case seeking a preliminary injunction to prevent the release of the software citing multiple exemptions under the WA Public Records Act including ensuring the security of public computer systems (RCW 42.56.420(4), an exemption for source code or object code received in the last 5 years (RCW 42.56.270(1), as a trade secret (RCW 42.56.270(11) and the Uniform Trade Secrets Act RCW 19.108 (King County Superior Court Case No. 18-2-19467-9 SEA). Insodoing, it sought claims

[41] Incidentally, the app was also based on work conducted at UW by researcher Traver looked up to named Karl Koscher; c.f. FareBot Github Repository, https://github.com/codebutler/farebot.

against not only the requestor Tyler Traver, but also Sound Transit-- as it is typical for a preliminary injunction to be filed against both the requestor and the disclosing agency.[42]

Vix declared that one of the two software applications requested, ORCA Inspection, is capable of re-valuing cards (i.e., both reading and writing information to the cards), which would undermine system and fare payment integrity if disclosed. In his declaration to the court as part of the injunction, Vix System Analyst Angelo Serraino attested to many of the same assertions in Vix's preliminary injunction:

> *"Both Applications constitute Vix intellectual property. Vix does not allow the public access to the Applications, and only allows third parties to use them pursuant to signed contracts. In fact, Sound Transit and the Agencies do not own the Applications, but instead have a license to use them only for the duration of Vix's contract. Further, only those employees who actually work on Vix software have access to source code" (Declaration of Angelo Serraino No. 18-2-19467-9 SEA).*

Here, Serraino clarifies that the agency does not have rights to the code, they only have a license to use it. He goes on to provide several reasons why disclosing the software would harm the vendor, the agencies, and the public; namely, that (i) it would be more challenging for the vendor to compete, (ii) bad actors could reverse engineer the system to harm system security, (iii) bad actors could falsify the values on transit cards using its read/write functionality, and (iv) the vendor's services would ultimately be more expensive for public agencies and taxpayers. Traver read Vix as asserting that the license does not grant the agencies the ability to inspect the software.

A notable part of Vix's argument was that disclosing the software would present a security risk. At the time we spoke, Traver had mistaken both of the applications to be read-only; the court filings and accompanying declarations of Vix assert that only one application (Officer Android) was limited to read-only. At the time, Traver told me of the information security concerns: "I think they brought [that] up because [security risks are] scarier to the court… [there would be] a legitimate security issue if it is capable of writing to the cards, but presumably that can be addressed by redaction or not giving me the key material" (interview March 19, 2019). Ultimately, however, the court decided not to draw a distinction between parts of the software that are security-sensitive and not so and granted a permanent injunction to withhold the software. In a settlement

[42] On February 12, 2020, Traver had requested the Settlement Agreement between Sound Transit and Vix which describes this sequence of events; it is available on the Muckrock website, titled "Vix Settlement Agreement," https://www.muckrock.com/foi/washington-54/vix-settlement-agreement-87992/.

between the two parties, Sound Transit also agreed to pay $5,000 to Vix—based on previous filings, this appears to have been sought to recover legal fees "to the maximum extent permitted under applicable law" (Summons August 6, 2018 in Vix Technology (USA) Inc. vs. Puget Sound Regional Transit Authority and requestor).

Before the case ultimately resulted in the court granting Vix a permanent injunction to withhold requested code, Traver's lawyer had found my related records requests exploring the boundaries of Vix's intellectual property on Muckrock.com and reached out to me. I recognized his lawyer from my interactions with her at the Washington Legislature Sunshine Committee.[43] She is a prominent member of the Washington Coalition for Open Government, the primary advocacy group in Washington organizing for the maintenance of a strongly pro-transparency Public Records Act. It struck me that during my engagement with the Sunshine Committee as a research assistant on the Transportation Data Collaborative, my stance was that the Act had concerning privacy implications for individuals and needed to be amended in ways that she vociferously disagreed with in our last interaction. As we discussed Traver's case, she learned of my personal view that the public should have some measure of accountability over proprietary software and asked me to write a declaration in his favor. At the time, I was interested to write the declaration, but did not follow through— ultimately I was ambivalent about being enrolled into the case study in the midst of my writing about it, and about the surprising turn that would have led me from occupying a critical stance on public records transparency to an (albeit distinct) advocacy role.

4.7 Conclusion

This case described the contracting relationship between a public agency (here, the ORCA partnering agencies) and their vendor Vix as they negotiate the boundaries of what data, hardware, and software belongs to the agency versus vendor. According to the contract, ORCA agencies owned data generated by the system, and the vendor owned the hardware and software it was

[43] "The Committee was created by Substitute House Bill 5435 in 2007 (codified in RCW 42.56.140). The bill established the Committee to review all public disclosure exemptions and make recommendations to the Legislature as to whether each exemption should be continued without modification, modified, scheduled for sunset review at a future date, or terminated. The Legislature stated that in light of the changing nature of information technology, record-keeping and the increasing number of public disclosure exemptions, periodic review of public disclosure exemptions is needed to determine if exemptions continue to serve the public interest." (Sunshine Committee 2019).

generated on. Over time the agency moved from a stance that contractually "owning" the data was necessary, even as the vendor managed the data warehouse, to a decision to "have" and warehouse their own longitudinal ORCA data (ironically through a cloud provider). While the original contract provided that the ORCA agencies would be able to access a version of the data that they owned at no expense, in practice, the technological frictions involved in extracting longitudinal data off of the "back office" system required the agencies in 2017 to sign an additional agreement with Vix and pay $285,182 in additional costs needed to export back office data. Around the same time, they began an effort to replace their existing vendor with one that would not only warehouse the data in the cloud, but also provide more modular, interoperable components to avoid future such technological frictions and lock-in. This part of the case study demonstrates that "owning" data is not always the same as having it or being able to access it without mediating factors.

This case also explored other parties' bids to access data, especially institutional customers of the ORCA card system who expected that by paying for ORCA cards, they should also be able to obtain data about their use. Notably, this institutional bid for data access was tied to paying for services-- but was not couched in the language of data ownership. While the WA Public Records Act initially granted institutional customers (i.e. employers) access to data about their employees' card use, privacy advocates pushed back until the Act prevented individual-level information from being released aside from exceptional circumstances. This section of the case study further underlines the close ties between data access and assertions of oversight and accountability, how the language of the Public Records Act draws distinctions between levels and forms of data access, and how the agency's interpretation of the Act itself plays a role in mediating the data institutional customers receive. Interestingly, this aspect of the case also exemplifies ongoing data access requests which are not asserted as claims to data ownership.

Finally, both the Public Records Act and claims of intellectual property also mediate what information individual residents are able to access. As I explore in the case of Tyler Traver and his requests for the software used by fare enforcement officers, for public agencies to "have" a copy of information (in this case, software used by ORCA Fare Enforcement Officers) does not ensure that the agencies own it. Public records requests like Traver's interrogate the boundaries of what information should be considered "public record," and the court process provides the opportunity to override the distinctions drawn between agency and vendor property as outlined in the vendor agreement. In this case, the court sided with Vix in issuing a permanent injunction from disclosing

the requested software, but in the process illuminated different parities' interests, through which core issues in this chapter of data ownership and sharing come into contact with larger questions of transparency, public trust, information security, competitiveness, and cost to taxpayers.

Chapter 5. CASE STUDY 2: THE KING COUNTY METRO-LYFT WOULD-BE PARTNERSHIP

The previous chapter explores a public agency's access to its vendor data. In this chapter, I investigate how agencies access data in a public-private partnership model. Specifically, I focus on a pilot partnership at King County Metro to subsidize a rideshare company expand service to public transit hubs, one result of a federal initiative to promote public-private collaboration called the Mobility on Demand Sandbox. I focus on the conflict that arose in negotiations with initial project partner Lyft around data sharing, and how these disagreements led the agency to replace Lyft with another company, Via, before the project could launch. The Washington Public Records Act played a key role in constraining access to rideshare company data, contributing to the delay and failure of the initial round of contracting. Here, data sharing conflicts illuminate frictions that persist even during a turn toward greater cross-sector collaboration advanced under the banner of "mobility on demand." I explore how the term "mobility" collapses distinctions between publicly and privately-operated transportation, and how programs to promote it (in their various forms) necessarily require greater cross-sector data sharing.

5.1 PUBLIC TRANSIT AGENCIES RESPOND TO THE ADVENT OF RIDESHARE

"Rideshare" companies provide app-based platforms to connect riders and drivers for hailing rides. Critics have noted that the pro-social valence of the term "rideshare" is a deliberate strategy to evade regulation under taxi and limousine oversight (c.f. Calo and Rosenblat 2017); for this reason the Associated Press Stylebook prescribes using the term "ride-hailing" instead (Warzel 2015; Zenner 2015). In this chapter I will use the term "rideshare" to be consistent with the term used in my field sites and forego the use of the term "transportation network company," which refers to the same business model and was also used interchangeably by my research participants.

Rideshare as a business model was created in 2009 with the founding of the company Uber; its popularity accelerated in 2012 as competitor Lyft entered the marketplace. Since its inception, rideshare has presented a unique set of problems for public transportation agencies; both Uber and Lyft have openly defied state and local law enforcement through lobbying against the designation of their firms as subject to the same laws as taxi and limousine services. Since its initial foray into

transportation services, Uber's business practice has been defined by its assertion that existing taxi and limousine regulations did not apply. The company was also found to have created a program called "Greyball" to evade law enforcement investigations, "The service used that tool to be able to operate in areas where it had no permission to do so, including Portland, Boston, Paris and Las Vegas" (Moon 2017). A 2017 New York Times article (Isaac 2017a) reported that the program included blocking local regulators from using the platform, tracking the SIM card numbers of phones available for purchase near police stations, and setting up the app to operate differently within geofencing around the offices of oversight authorities. Importantly, the Greyball case is indicative of the challenge public agencies across the United States have faced in getting access to rideshare companies' data; in this instance, the risk of the transit agency oversight merited Uber's additional expense to circumvent this data collection.

As the popularity of rideshare gained critical mass, public transportation agencies became increasingly interested in collecting data about rideshare operations. The primary purpose of data collection was to measure the impact of rideshare usage on local transportation patterns. For example, data on rideshare use to and from crowded airports and stadiums could provide meaningful insight into the potential impact of designated pick-up and drop-off locations on congestion (Lin et al. 2016). While rideshare companies marketed themselves as an environmentally friendly alternative to individual vehicle ownership, one which would replace individual ownership long-term, agencies had pressing questions about the net impact of additional rideshare company vehicles on local transportation infrastructure (Erhardt et al. 2019), particularly given that the companies had no incentive to limit the number of drivers signing on to work.[44] At stake were larger questions about how rideshare might be regulated to ensure high safety (Feeney 2015; O'Brien et al. 2018), environmental (Henao and Marshall 2019; Erhardt et al. 2019), and labor standards (Calo and Rosenblat 2017; Li et al. 2019).

In response to sustained pressure from both municipal governments and the research community over their lack of data sharing, in early 2017 Uber released a dashboard for accessing anonymized data from billions of trips called Uber Movement "to help urban planning around the

[44] Since rideshare companies' initial foray into the market, city governments have passed some laws that apply to rideshare companies in the face of vigorous industry pushback. To date, city governments have explored or passed regulations that rideshare companies allow their drivers to unionize (Isaac, Wingfield, and Scheiber 2015; Hawkins 2017), cap the number of drivers on the road (Marshall 2019; Li et al. 2019; Lawler 2014), mandate background checks (Mehta 2016), cap time spent cruising (Marshall 2019), or provide handicap accessible cars (Mapelli 2017; Casey 2017).

world."[45] Signaling the beginning of a shift toward collaboration, a newspaper article at the time remarked that "Uber [was] hoping to convince cities it can be a cooperative player in this new world of ride-sharing and multimodal transportation. But that will be difficult, given its history of tussling with officials over regulations and licensing" (Hawkins 2019). While most news coverage presented the dashboard as an "olive branch" (Isaac 2017b), this coverage also hints at the limited usefulness of proactively disclosed data for planning purposes: "City officials said that they appreciated user data privacy concerns but that they also hoped to see more useful information from Uber. [A transportation planner] shared a list of detailed requests that could aid future urban development, like demand patterns around car-free tenant housing, locations with likely potholes and the most common pickup and drop-off locations" (Isaac 2017b). During my fieldwork, my participants viewed the Uber Movement dashboard with skepticism, as a cynical means to forestall more useful data from being shared.

The programs I describe in this chapter trace a critical moment in which the adversarial relationship between regulators and rideshare companies gave way to an exploration of new possibilities for public-private partnership. In the next section, I describe a pilot program between Lyft and transit agencies in King County and Los Angeles County, and how conflicts over data ultimately led to the abandonment of their original partnership agreement with Lyft in favor of a partner called Via more amenable to sharing. During this same time period, a noteworthy court case between Lyft and Uber and the City of Seattle ruled that claims of trade secret were not enough to bar access to data. This case highlights the role of data sharing in cross-sector partnerships and its entanglement with larger questions of accountability, oversight, and regulation. I focus especially on assertions of "data ownership" made by different parties as bids to gain or bar access to data in these discussions.

5.2 "Mobility on Demand"

As rideshare service usage increased over time, public transportation agencies began investing in programs to re-envision their work to include them. As one transportation planner I spoke to explained:

[45] Uber Movement Tool https://movement.uber.com/?lang=en-US.

> "As [rideshare companies] start to gain demand and they become a solution that everyone (not only transportation wonks) becomes familiar with, then the question becomes, 'What can they do? And that's when the wonks themselves —ourselves—engage. What is the potential here to address some of our classic problems? ... [How rideshare companies] might have an opportunity to improve transit services?" (interview April 16, 2018).

Agencies in many jurisdictions began to refer to these initiatives as "mobility on demand," a term that arose in transportation planning around the same timeframe that I report on this chapter. This shift in focus was also evident in the language agencies used to describe their work as providing "mobility" or "mobility as a service" instead of transit (Kamargianni et al. 2016), and to describe individuals they served as "consumers" instead of riders.[46] Mobility on demand highlights real-time connectivity and innovation, it also builds on prior trends through an emphasis on contracting with private firms. A 2019 report from the venerable Transportation Research Board highlights how the concept commoditizes transportation:

> *"[Mobility on demand is] an innovative transportation concept evolving around connected travelers, where consumers can access mobility and goods delivery services on demand by dispatching or using public transportation, shared mobility … and other innovative and emerging technologies. The U.S. Department of Transportation (DOT) defines mobility on demand as a concept based on the principle that transportation is a commodity where modes have economic values that are distinguishable in terms of cost, journey time, wait time, number of connections, convenience, and other attributes." (Shaheen et al. 2019 emphasis added).*

While the movement from transit agencies to explore partnership with rideshare companies was novel, this phenomenon joined a much longer history of public agencies attempting to cut costs and promote efficiency through contracting (E. D. Sclar, Schaeffer, and Brandwein 1989; Girth et al. 2012). Importantly, the mobility paradigm asks transportation agencies to support consumers in making individual choices about transportation, of which transit is one option (Shaheen et al. 2019). This "mobility on demand" paradigm is distinct from a longer lineage of contracting in its emphasis on frictionless interoperability between public and private transportation options,

[46] For instance, the 2017 City of Seattle Department of Transportation New Mobility Playbook refers to public transit as one of several transportation modes, collectively referred to as "shared mobility services" in a way that diminishes the distinction between public and private services. "More recently, shared mobility services (including public transit), real-time travel information, and other digital technologies are providing "à la carte" mobility and customer experience offerings that get people where they want to go based on their needs. These new ways that people interact with transportation infrastructure are at the heart of what we refer to as the new mobility." (Seattle Department of Transportation 2017)

payment, and information technologies (Kamargianni et al. 2016), which necessitates a greater degree of information sharing.

This vision found expression in ideas for new apps, partnerships, and programs. For instance, in 2017 in the City of Seattle, the head of the Department Transportation Scott Kubly tasked his team with exploring how to create a "mobility on demand" app to provide information to consumers on different transportation modes' cost, journey time, and convenience. While the proposed app's data sharing needs and costs ultimately proved too great for the department to bring into reality, the imaginary of frictionless, ready-at-hand information exemplifies of the promise of this vision to transportation planners; namely, that through translating different transportation choices into information, such as price signals, individuals can be incentivized (for example, through subsidies) to adopt behaviors that better adapt transportation infrastructure to demand, and individual choices to environmental benefit (e.g. Griffin et al. 2016).

This mobility on demand vision included an intention for robust data sharing to allow for inter-modal transfers, real-time information, and cost-sharing. As federal and local transportation agencies had begun to explore new ways to expand transportation access, they sought to support an array of transportation modes, such as rideshare and bikeshare. With these new efforts came unanswered questions about how data sharing would work in practice (Grossman and Lewis 2020). In the next section, I demonstrate how funding was structured to support experimentation and sharing out lessons learned.

5.2.1 *Federal funding for the Mobility on Demand Sandbox program*

Federal funding followed the mobility on demand vision. In 2016, the Federal Transit Administration, an agency within the U.S. Department of Transportation issued an $8 million call for mobility on demand proposals for programs that would advance the vision of a "a multimodal, integrated, automated, accessible, and connected transportation system in which personalized mobility is a key feature" (Federal Transit Administration 2019). A key intention of this program was to foster fresh thinking, as the federal agency also supports state and local governments with the financing, planning, and implementation of public transportation projects. Originally known as the Urban Mass Transportation Administration until 1991, underlining its founding focus on

mass public transit, the federal agency also plays a role in facilitating learning between large metropolitan areas.

It is in this role as a funder and facilitator that the Federal Transit Administration issued an $8 million call for proposals to find the first phase of a long-term exploration for how transportation agencies could better expand service access by integrating and collaborating with rideshare firms like Lyft and Uber. The resulting 2016 funding call for proposals known as the Mobility on Demand Sandbox was intended to fund "bold and innovative" pilot partnerships between transit agencies and rideshare companies to expand transit access in grantee metropolitan areas across the country. Significantly, this effort was led from within the Federal Transportation Administration's Office of Research, Demonstration and Innovation; the call for proposals encouraged projects with a research dimension to evaluate their effectiveness. The funding aimed to deploy and test different partnership models, such that projects would be able learn best practices from each other before the next iteration of funding. The research ethos of the project was reflected in its program objectives, which sought to both assist agencies in developing the ability to integrate mobility on demand practices with existing services and to validate its feasibility as a business model. The funding also sought to measure the impact of mobility on demand programs on residents and transportation systems.

Project selections were announced in October 2016. Among the 11 grantee programs announced was a pilot led by the Los Angeles County Metropolitan Transportation Authority, known as "LA Metro," the county-wide transportation agency that operates public transit in the Los Angeles metropolitan area, jointly with Sound Transit and the King County Metro Transit Department, known as "King County Metro," the county-wide transportation agency serving the Seattle metropolitan area.[47] The two agencies received $1,350,000 for a pilot project which each agency would create similar programs to compare outcomes. The rideshare company Lyft agreed to provide discounted rides to and from three transit stations in each metro system. Notably, this effort was led by offices for innovation in both transit agencies—an office formally known as Office for Extraordinary Innovation at LA Metro, created in 2015 "to identify the best ideas in transportation and help to test, refine, and implement them at LA Metro" (Office of Extraordinary

[47] Sound Transit was a key partner in the Seattle area, necessary to ensure interoperability with the ORCA payment card. Given that King County Metro and Sound Transit worked in tandem on this project, and because of the additional complexity introduced by the partnership between King County Metro and LA Metro, in this chapter I refer synecdochally to King County Metro instead of both King County Metro and Sound Transit.

Innovation, n.d.) and at King County Metro. This pilot program was intended to improve access to public transportation in less densely populated areas, where transit users often live farther from the nearest transportation access point. A key lead on the project managed King County Metro's Innovative Mobility Program, which was tasked with "exploring, testing, piloting, and doing research about the ways in which emerging mobility services and new technologies can move [people] around in King County" (interview April 16 2017).

5.2.2 *"First mile last mile"*

In transportation planning, efforts to increase riders' ability to more easily get to and from the transit system are referred to as "first- mile last-mile." These efforts are especially salient in planning transit for residents of a large metropolitan area where there is less housing density; long-term, the vision for first-mile last-mile transit is to decrease the number of single occupancy vehicles needed, and to increase short trips to and from high throughput transit hubs. "Park and ride" hubs for access to transit from these areas were at capacity; providing additional transit to and from the hubs held promise as a cost-effective alternative to adding additional parking.[48] In their joint proposal for the 2016 funding from the Federal Transit Administration, Lyft signed onto a non-binding letter of intent with LA Metro and King County Metro, agreeing to be the transportation provider to and from transit hubs to support commuter access to light rail and bus transit. In exchange, LA Metro and King County Metro would subsidize each ride to encourage residents to adopt the program. At stake was the question of how transportation agencies would respond to the rise of rideshare companies as a proportion of transportation volume.

5.3 DATA SHARING NEEDS

Data sharing was an important part of enabling Los Angeles Metro and King County's Mobility on Demand program for two reasons. First, both agencies needed visibility into whether each subsidized trip went on to use public transit. Second, agencies wanted to integrate the payments they made to subsidize qualifying trips with the app used to hail the rideshare. In the

[48] At the time, this idea was one idea that King County Metro was still exploring; other ideas included improving bike path connectivity to park and ride hubs.

initial non-binding letter of agreement that Lyft signed to support the funding proposal, the company agreed to collect and share data with each agency to enable these oversight and payment integration functions. Specifically, agencies asked Lyft to deliver data on each trip, including:

- trip-level ID numbers,
- driver ID numbers,
- prices charged,
- routes taken,
- cost and fee for the trip, and
- ORCA card numbers of riders.

This information would assist in both transfer integration and for auditors to determine whether the subsidized trip qualified for a subsidy—by transferring to or from a public transit commute. This plan also would require the company to integrate transit card information with its own application to enable payments and to verify transfers.

Data sharing was also necessary to evaluate the program. Researchers at the University of California Berkeley and the University of Washington had signed on to the initial proposal to evaluate program outcomes in Los Angeles and King counties respectively. At UC Berkeley, a team led by Susan Shaheen, the co-director of the Transportation Sustainability Research Center and the Director of Innovative Mobility Research. At UW, the project was led at the Washington State Transportation Center by Mark Hallenbeck. Both Susan and Mark had years of experience studying public agency transit data but rideshare companies to date had long withheld sharing with local transit agencies and researchers. This project presented a unique opportunity to researchers to work with rideshare companies to understand the impact of their services on the transportation system.

The research planned to focus on how the partnership would impact transit agencies' mission of promoting "affordable, equitable, and accessible" transit access (King County Metro 2017). Specifically, the researchers planned to evaluate how the program expands transit ridership with respect to age, ability, income, English proficiency, and people of color. They also planned to consider the program's ability to divert parking use in transit part and ride stations that were at capacity. To answer these questions, the researchers would rely on Lyft to share data with their team, as well as on additional stakeholder interviews, surveys with riders, and usage data with a focus on the demographic and socioeconomic dimensions of ridership.

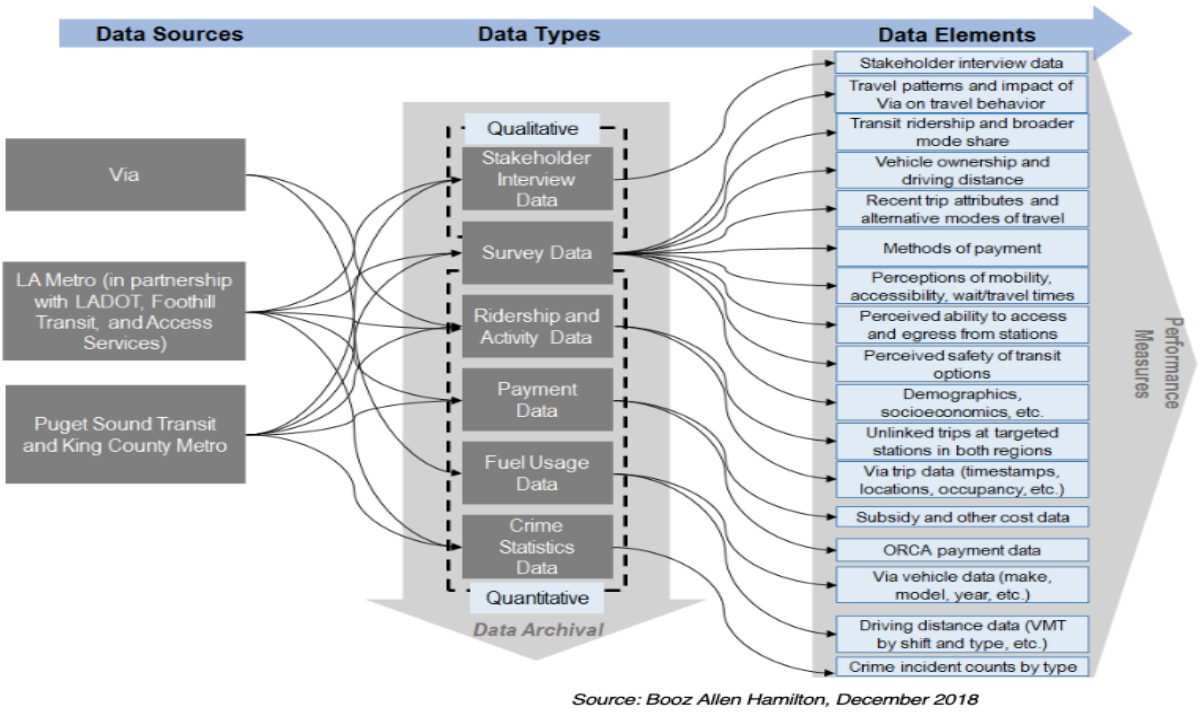

Figure 9 This figure describes the data sources that will be used in MOD Sandbox Program Evaluation by the Berkeley and UW research teams. The figure refers to the agencies' replacement partner, Via.

5.4 Contracting Frictions

In order to apply for the Mobility on Demand Sandbox grant, the partners signed a letter of intent outlining key tenets of the agreement and the intent to participate. The Federal Transit Administration widely publicized the grant awardee partners, naming Lyft as part of the bid to provide the service. These press releases summarize the content of Lyft's early commitment to "first/last mile solutions for trips originating and ending at select transit stops. Customers [were to be] able to use a mobile app or call a dispatcher phone number, providing equity to individuals who are either unbanked or non-smartphone users" (Shared Use Mobility Center, 2017). This agreement included a verbal agreement to supply all the data necessary to enable and evaluate the program (Grossman and Lewis 2020). After the announcement in October 2016 that the Federal Transportation Administration would be funding the partnership, the project timeline allowed one year to formalize and stand up its services.

While the initial timeline for this project was to launch the pilot in 2018, negotiations had stalled the projects for many months. The project had hit delays as it moved from the initial letter of intent into the pilot. As the Innovation Mobility Program Manager described it; “almost to the summer of 2017, we were not [making] any kind of tangible progress with Lyft…There was never agreement with Lyft about the scope of the project and what it was and who is responsible for what, and again we thought we had that agreement in the application process and when they signed on to that application-- and I guess we came to realize maybe there were different understandings of what was written meant” (interview April 16 2018). Adding to the complexity of these discussions was the number of parties; the pilot was intended to allow for evaluation across both LA County and King County, with multiple transit agencies participating in each site. These frictions ultimately contributed to the dissolution of the agreement between Lyft, LA Metro, and King County Metro**.** Key areas of disagreement included “differences with respect to data sharing, provision of multi-lingual call center support [for equity in access], and delivery of wheelchair accessible vehicle service as a part of service delivery” (Grossman and Lewis 2020). While accessibility concerns ultimately contributed to the dissolution of the partnership, a researcher evaluating the Mobility on Demand Sandbox project who was close to these negotiations explained that data sharing concerns were very salient: "They had agreed to a certain level of data sharing as part of the pilot, [but] as they scoped out the project and worked on the deliverables, Lyft wasn't comfortable with the level of data they wanted" (interview, January 20, 2019). While there were several factors that led to the failure of the contract with Lyft, this respondent highlighted Lyft’s reluctance to share as much data as had been initially agreed upon as a primary factor.

Data access in particular was described to me by King County Metro as a “non-negotiable.” Nevertheless, the program required the agency to work with “these big conglomerates, and they are trying to figure out also how much they can give, and they too have their non-negotiables. And if you ask me, they don't make sense some of their non-negotiables, right?” (interview April 16 2018). During the discussions formalizing the data needed, Lyft also signaled reticence to share data with the research university partners that would be evaluating it. As key personnel on the university research team described to me, “The bid was a year ago. The active discussions have been 6 months so far. They had not thought through that we would need data. Lyft has already signed on to the mobility on demand project. We know we have to solve this sometime” (interview April 26, 2017).

5.4.1 *Data sharing tensions and breakdown*

Not only was data sharing essential to enabling and evaluating the program, it would be necessary to subtend the plans to pay Lyft as a service provider. However, after the initial non-binding agreement was signed that led to the project being funded, it became clear over the course of the planning discussions necessary to stand up the program that the data that agencies asked Lyft to provide (as well as the content of the prices they charged customers) were considered sensitive information by the company. As one respondent explained to me at the time, "The algorithm that determines their trip price is a carefully guarded business secret. The price is based on the likelihood of your ability to find a ride and other factors. At this point, [the agencies] are struggling to decide what the subsidy value will be, and they have a fixed budget to work in" (interview TBA MH). This quote highlights the difficulties the agencies had in setting an amount they would agree to pay to subsidize each trip in the absence of more complete pricing information about how much trips would cost the end-user. While the initial plan involved the agency subsidizing a share of an on-demand Lyft ride, but the initial agreement did not specify how the subsidy would work. Further complicating this analysis was the agencies' interest in identifying a level of subsidy that would incentivize drivers to forego single occupancy vehicles in favor of transit rides.

Disagreements over data sharing between Lyft and King County Metro became a primary focal point for the university-based data sharing effort I was working on as a Research Assistant. Understanding different parties' data sharing needs and concerns was part of my job at the time, in particular because we wanted to create a data sharing repository to bridge the gap between these different parties' stated needs. From this vantage point, I encountered multiple direct and indirect stakeholders over the course of these negotiations and came to appreciate the content of each party's conflicting expectations and rationales. At the time, the team behind our data repository (the UW Transportation Data Collaborative) considered the data too privacy-sensitive for agencies to collect and manage directly, given the risks of privacy breaches in the event of a public records request. Although the details of our discussions with rideshare companies were confidential, broadly they considered the UW effort, as an intermediary, to be a very promising alternative to any scenario in which agencies required data be disclosed to them directly.

5.4.2 *King County asserted data ownership rights*

King County Metro adopted a strong stance in favor not only of receiving access to the data collected by virtue of the existence of the program but asserting ownership rights over the resultant data.

> *"Let's go back to some of the key issues, we have the data question, right, so one of our foundational principles to our pilot is that we need to get the data, right? We understand that say an Uber or a Lyft doesn't want to share their current data with us, so, fine. But, if we do this pilot, one of the main reasons is to actually have the data, right, it's not just to kind of play with the pilot and see it on the ground, and it's not to get aggregate metrics. It is to actually have the data, to be able to understand it, to be able to understand what we did…So, to me if we are paying for the rides, then we own that data as much as anyone else. And so we can get into a discussion about protecting the data, right, from privacy disclosures, be it for a [personally identifying information] perspective or for their business interests, right, but don't tell me that I cannot have—cannot see that data in perpetuity." (interview, April 16 2018, emphasis added).*

Notably, he attributes that ownership to the virtue of who is paying for the service. Later in the interview in discussing the breakdown in the Mobility on Demand negotiations, he reprised the agency's position as "Of course you should give us our data, we are paying for the service, we should be able to do whatever we want to it." Note his reference to the data as belonging to the public agency ("our data.")

The Innovation Mobility Program Manager spoke candidly as to how the breakdown in discussions with Lyft had informed a successive call for proposals for a new first-mile last mile program. This program was explicit from the request for information stage that it would require granular primary data for the purposes of auditing and oversight. Describing the lessons learned from two failed contracting sessions with Lyft (and a previous first-mile last-mile effort that stalled with Uber), the Manager recounted:

> *"From this run of projects, we ended up having no project with at least a classic big 2 [rideshare company], so I think there's something to be said about working with them in this space as well [For our new project] we said we need primary, raw data, and actually if you look at the RFP, it says something like if for some reason you feel that you cannot just share that data with us, then you the provider need to explain the ways in which you create the platforms for sharing that data with us, and the cost of that, you have to make*

> *that part of the proposal… [We got] the sense that the [rideshare companies] in their Spring 2017 iteration at least, it wasn't clear whether they could provide all of that" (interview April 16, 2018).*[49]

This passage distills the agency's lessons learned about rideshare companies' reservations toward providing primary data, and how it incorporated that learning into their next wave of contracting to successive programs. He goes on to explain that the agency had learned to look "beyond" rideshare companies to other business models in the new mobility space.

5.4.3 *LA Metro asserted data ownership rights*

LA Metro mirrored King County Metro in its strong assertion of the agency's rights not only to receive data, but to own it. At LA Metro's Office of Extraordinary Innovation (OEI), issues of data sharing have been "very connected" to conversations about data ownership. As a transportation planner in the office closely involved with the Mobility on Demand Sandbox initiative explained:

> *"[Our office] has a few different perspectives on the issue of intellectual property and what it means to pay for [services]... one is that if a public agency is paying a private entity to make a bespoke service for that agency, that agency should be able to own the intellectual property that comes out of whatever is learned from that project. By nature of doing business with the public agency, the vendor is learning a lot. Another perspective is that if a vendor is doing business with a public agency, making modifications to their product, there should be a licensing agreement for the use of that information...[so] you can use it" (LA Metro Transportation Planner, interview August 8 2018).*

The respondent went on to characterize the Office as ahead of other agencies in the U.S. with respect to the maturity of its thinking on data ownership and rights.

---

[49] Note that the respondent refers to platforms as a solution when providers "cannot just share that data with us"—an allusion to dashboards like Uber's Movement as well as third-party intermediaries like the UW Transportation Data Collaborative, which had been actively pitching to both King County and LA Metro over the prior year. In our interview, he also referred directly to the UW Transportation Data Collaborative (knowing me through my work on that project) to contrast his views on data ownership and access with the views advanced by the Collaborative, which was among other purposes intended to mediate agency data requests:

> "So I guess you're familiar with this discussion, but it goes beyond-- and why for instance once again, my differences with the Transportation Data Collaborative, or why I cannot wrap my head around it, is because I think--and knowing--and having dealt with a couple of these contract negotiations that there is *no incentive whatsoever* for a [rideshare company] to allow for any additional queries within the data that is [shared] … my questions today are my questions today. My questions in [1 or 2 years from now will] be building on my experience … I might be looking at what other people have done, and so I just cannot-- I just don't think I'll ever have all the quesitions right now" (interview April 16 2018).

Even in the face of ongoing pushback from Lyft, both LA Metro and King County Metro remained committed to the requirement for direct data access to requested data generated in the provision of the service. One aspect of this argument was based on principle; both agencies stated their conviction that because the data was being generated as a result of a publicly subsidized service, agencies had a rightful claim to resulting data. As an employee of LA Metro explained, "The rationale behind the program is serving the public interest. You need a [rideshare company] willing to work with a public agency, that willingness would require being comfortable with a certain level of data sharing - if we are using public dollars to pay or subsidize rides, that is coming from a public agency and has obligations to taxpayers. There is a certain level of data we expect" (interview August 8, 2018). Another part of the argument in favor of data sharing was pragmatic—certain aspects of the program such as interoperability, auditing, and research would require robust data access. Data sharing was necessary both as a primary input to enable the program, and a secondary driver of its evaluation. Ultimately Lyft "backed away" from involvement in both LA Metro and KCM's MOD Sandbox efforts in 2017.

5.4.4 *Lyft's concerns about data disclosure under the WA Public Records Act*

Throughout these negotiations, Lyft was reticent to share data with both King County Metro and LA Metro commensurate with the amount needed to support the pilot. In other municipalities, data about Uber and Lyft's ridership and driver compensation was used as a starting point for regulatory intervention. During the same timeframe that the Mobility on Demand program was being formalized and Lyft was making decisions about the level of data to share, it was simultaneously entangled in a lawsuit with the City of Seattle over a public records request. Both Lyft and Uber were appealing a pending records request to the City for it to disclose data that had provided by the rideshare companies as an obligatory part of the City's permit to operate in city limits. Lyft's reticence to work with LA Metro and King County Metro had been informed by this experience of their perceived data exposure and ongoing uncertainty. Both companies asserted that both privacy and competitiveness harms would result from their data being disclosed in the event of a public records request. Others who spoke to me close to these negotiations indicated that they were primarily concerned about transparency in data sharing giving rise to further regulations.

From the perspective of my work with the UW Transportation Collaborative, Lyft's stated concerns about data exposure were consistent with our objective of finding additional public agencies and private firms willing to sign on to use our intermediary; Prof. Jan Whittington and Mark Hallenbeck had been working to raise awareness among public agencies about the potential risks of retaining sensitive locational data, such as its high likelihood of re-identifiability. In this respect, our team promoted a vision distinctly different than that of King County Metro or LA Metro, namely, that firms should not be asked to provide sensitive data directly to agencies in light of the high-transparency public records context. As King County Metro struggled to achieve consensus with Lyft, it contemplated whether to work with our effort, given the firm's preliminary enthusiasm and endorsement of our model.

5.5 Public Records and Rideshare Data Sharing

The most effective data access measures local governments have adopted are the product of making rideshare companies' permit to operate contingent on some level of data sharing. In 2014, the City of Seattle had revised Ordinance 214524 on transportation network companies (like Lyft and Uber) and taxis to make its permit to operate contingent on sharing zip-code level aggregated data, in order to get a sense of how many vehicles and rides were operating in the city.[50] Data sharing has been a compulsory part of the permitting process for granting rideshare companies permits in many cities; for example, the New York City Taxi and Limousine Commission requires rideshare companies to satisfy granular reporting requirements, such as how many cars are on the road. New York's success in mandating data sharing has been attributed to the size of its market---firms' revenue would be significantly impacted if the City of New York did not grant them a license to operate.[51] Smaller markets like Seattle (with more disclosive public records laws) have had more difficulties gaining access to firm data, facing pushback from firms on the scope and granularity of data to be shared given the fact that many cities used data about rideshare companies to establish protections for drivers, and other regulatory objectives. A 2014 initiative from the

[50] http://clerk.seattle.gov/search/ordinances/124524

[51] The New York City Taxi and Limousine Commission has an unusual arrangement in that the state public records law in the State of New York, the Freedom of Information Law, enables the state to collect more information. The City used this data to investigate and then support a 2018 law intended to ensure that rideshare drivers earn a living wage.

Seattle City Council attempted to limit rideshare companies to 150 drivers each, in an effort to protect would-be drivers from joining the platform in a phase of rapid expansion---risking the possibility that a glut of driver supply and insufficient demand would make it difficult for any drivers to cover their operating costs. In "a successful petition drive backed by the companies [to undo these measures] led to a deal in which the driver limits were scrapped and, in exchange, the companies agreed to submit quarterly reports on how many rides they were giving, what ZIP codes the rides were serving and other data" (Gutman 2018).

Two years after Seattle began to make permits contingent on data sharing, in 2016, a man from Austin, Texas had filed a public records request for data about Uber and Lyft's operations in Seattle. As mentioned in Chapter 2, public records requests in Washington are made directly to public agencies; where requested data concerns private companies, the agencies provide the firms with "third-party notice," a courtesy letter issued in advance of disclosing the data, to give firms the opportunity to file an injunction. The requestor Jeff Kirk was an Austin resident who had been involved with local efforts to regulate rideshare and had been frustrated at the companies' lack of transparency in Texas. When he learned that Washington law allows for even those living outside the state to submit a public disclosure request, he filed for the most recent two quarters of data provided to the City of Seattle on:

- The total number of rides provided by each rideshare company.
- The percentage or number of rides picked up in each ZIP code.
- The pick-up and drop-off ZIP codes of each ride.
- The number of rides when an accessible vehicle was requested.

Kirk told me:

> *"I was not requesting location traces beyond anonymized ZIP code data, if only because I knew there was no chance I could obtain it even if I wanted the data -- which I did not. (In a nutshell, I concur with [Uber's and Lyft]'s view that providing such a granular level of data raises a considerable number of privacy concerns, particularly if said data were to be made available under PRA request." (personal communication, September 28, 2017).*

In this quote, Kirk undermines the rideshare companies' argument that the data he requested could harm riders' privacy and highlights his intent in only gaining access to safer, aggregated and anonymized versions of data.

In response, the Seattle Department of Transportation provided third party notice to Lyft and Uber in notifying both firms that there had been a request for records pertaining to information they might consider to be proprietary. Lyft and Uber filed a third-party injunction to stop the disclosure on the grounds of trade secret, citing the fact that because there are only 2 firms in the market, releasing aggregate data still implicates firm competitiveness.[52] The court granted the injunction, barring the agency from disclosing the records until the matter could be decided in court.

Lawyers for Uber and Lyft argued that the requested records, although aggregated to the zip code level, contained data that would threaten the competitiveness of their businesses--especially with respect to information that each firm hoped to protect from the other--its closest competitor in the marketplace. Those writing in favor of releasing the data highlighted how it could be used to assess the racial equity of who is served by rideshare through analyzing service provisions across zip codes (Gutman 2018). Given the United States' history of redlining, zip codes are highly correlated with race (Bruckner 2018).

After multiple stages of appeals, the case was elevated to the Washington State Supreme Court. The judge, re-stating the language of the exemption for trade secrets in the event that "disclosure would clearly not be in the public interest and would substantially and irreparably damage a person or a vital government interest" found instead that disclosure of Lyft and Uber's zip code level data was in the public interest. In her decision, Justice Debra Stevens refers to the importance of data access by researchers in that it provides the ability to reveal redlining in rideshare. She cited a study that demonstrated that Black people in Seattle experience longer wait times on the Uber platform (Gutman 2018). Her decision also underlined the idea that government is not the only party with the right to conduct this oversight– but that researchers too should have access to proprietary data if it is in the public interest. Uber and Lyft lost this case in 2018, midway through the discussions with King County Metro and LA Metro I describe above and went on to cite the case as evidence that any data they shared with transportation agencies in Washington

[52] Careful readers will observe that the firms did not request third party injunction on the grounds of riders' privacy harms; note that the way the Washington Public Records Act defines an invasion privacy as information that would be *both* "(1) Would be highly offensive to a reasonable person, and (2) is not of legitimate concern to the public;" this definition is constructed purposefully narrowly so as to make invasion of a privacy a more difficult rationale for exemption (RCW 42.56.050).

State would be liable to be requested. Ultimately, this court decision demonstrated that data ownership, here asserted in terms of trade secret, was not sufficient to bar public access to data.

Publicly, both Lyft and Uber claimed to be most concerned about the privacy and competitiveness consequences of disclosure under the WA Public Records Act. In this context, competitiveness refers to both Lyft and Uber's stated reluctance to share data that would be published in aggregate; because the rideshare marketplace is dominated by two firms, they stated they were concerned that aggregated data segmented by zip code is easily "reverse engineered" by their primary competitor, which would know its own activity in that geographic area. To the extent that one company's performance in a zip code is related to the way they dispatch drivers across city space, they claimed that this data revealed too much about their market strategy to their rivals. Despite these public concerns about their rivals and the privacy of their riders, privately they were also said to be most concerned about the regulatory consequences of their data exposure to government. These represented a distinct set of concerns; while neither Lyft nor Uber has publicly stated that their data sharing concerns are related to the potential for further regulation nor the precise forms of regulation that most concern them, other marketplaces can be instructive. For instance, in New York City, the state Freedom of Information Law enables local regulators in the Taxi and Limousine Commission to mandate the disclosure of granular data which they later used to enact the first driver minimum wage in the county and a cap on the limit of drivers that the rideshare companies are allowed to sign on to their platform (remarks in a December 5 2018 event at Data & Society by Rodney Stiles of the NYC Taxi and Limousine Commission). Lyft and Uber fiercely opposed these labor regulations in New York City on the grounds that they would raise costs for riders and decrease demand, ultimately affecting drivers' take-home pay. More than a year after these measures were put into place, the New York City wage floors were still in effect. One interviewee, at the time an employee of the City of Seattle Department of Transportation, reported that behind the scenes, Lyft explained to him that:

> *"Their lawyers are saying, 'If we volunteer this information, we expose ourselves across a variety of regulatory environments.' We were going around and around. And they said it's not competitive risk that we are worried about, it's regulatory risk; we are worried about you seeing that the service does not do X or Y, does not serve a certain population. I only bring this up to suggest that they may not be as collaborative as they appear," (interview April 26, 2017).*

In this quote, the respondent confides that while the public policy personnel of rideshare companies were supportive of increased data sharing, the same firms' legal counsel advised against data sharing given a sense that it would enable new forms of scrutiny and intervention from the state into the rideshare firms; business models. As our data repository effort began to pursue rideshare companies as possible affiliate members, public employees would share their experiences with protracted negotiations and concerns about whether for rideshare firms, delay was an end in itself. Notably, this intimation undermines rideshare companies' argument that the Public Records Act itself was a barrier to data sharing and supports an alternative view that perhaps any degree of exposure to agencies themselves was a legal concern.

Back in King County, Metro employees also sensed that Lyft may be at cross-purposes with itself; that is, its reticence to share data was not consistent with their initial interest in government partnership. As key personnel who had negotiated both with Lyft and a previous stalled project with Uber described it, "If you ask me, they don't make sense some of their non-negotiables, right? If you are saying that [first-mile last-mile] is what you are trying to do" (interview April 16 2018). He went on to describe how he felt that his counterparts at Lyft were being overridden from higher levels within the company, especially by risk averse legal teams.[53] Moreover, the dollar amounts attached to the pilot public sector partnerships were perhaps not enough to merit the perceived risk of sharing the data, which he referred to as a vulnerability: "Sometimes I felt with the [rideshare companies] that these projects are so marginal to the bigger kind of picture of their worth, that why would they give the house, why would they allow the chink in their armor … for example in the case of [the failed talks with Uber], it was like at best $50,000" (interview April 16, 2018). Here, the respondent asks openly why rideshare companies would subject themselves to the regulatory risk of greater transparency to local transit authorities in the form of data sharing for the seemingly small income afforded by the pilot programs. The metaphor of "armor" underlines the sense in which rideshare data was walled off to agencies, steeled from view.

[53] Meanwhile, the City of Seattle Department of Transportation was facing similar challenges in eliciting data from rideshare companies for their own mobility on demand initiatives and shared the view that these programs were being stalled due to perception of risk. As one person who worked at the Seattle Department of Transportation shared, "[Lyft's policy personnel] get told no by the lawyers; so we are trying to go to the top. The whole nation is trying to do this [data sharing between rideshare companies and agencies]. The management issues are not unique. If we solve this, we have a huge market for them and for us" (interview April 26 2017). At the time, the City of Seattle was open to working with UW Transportation Data Collaborative as a data sharing intermediary, in contrast to King County and LA Metro's conviction that the data would belong to the public agencies as a matter of principle.

5.5.1 *LA Metro and King County Metro find a new partner*

Key disagreements between the partnering agencies and Lyft surrounded both data sharing and equity issues, namely the need to support rider accessibility and paratransit and call centers for riders who do not have cell phones. These contentions led the partnering agencies to select a different provider. By early 2018, both agencies had pivoted to sign instead with Via, a much smaller transportation service network provider which had a business model based on providing fleet vehicles to drivers (rather than rideshare firms, which relied on drivers to own their own vehicles). A primary determinant of choosing Via as a provider was that they could support paratransit and accessible vehicles, central to the transit agencies' mission and a requirement of Federal Transit Administration funding. The further contract negotiations between the agencies and Via required additional agreement on where to place the transit hubs, what the cost structure would be, what data would be collected and shared, how to provide wheelchair accessible service and call centers, and how the app would be designed. Whereas both agencies had originally planned to launch their MOD Sandbox services in July 2018, after pivoting to work with Via, LA Metro ultimately launched in January 2019, while King County launched in April 2019.

5.6 Framing Transportation Planning as "Mobility"

As I previously discussed, the term "mobility" instead diminishes the salience of the distinction between publicly provisioned and privately provisioned transportation. The projects, offices, and initiatives embracing "mobility" evince a shift in agencies' thinking about how to manage public transit services to managing urban movement overall. In pursuit of this goal, transportation planners in the agency began investigating ways that rideshare companies can play a role in "help[ing] improve transit services" (Ibid.). Other terms like *shared use* vehicles de-emphasized whether a service was publicly or privately operated, while also evoking the "sharing economy."

This vision of mobility was evident at the annual conference of the Shared Use Mobility Center (SUMC), called the Shared Mobility Summit.[54] I was invited to attend the event to represent

[54] The Shared-Use Mobility Center defines shared mobility as "transportation services and resources that are shared among users, either concurrently or one after another. This includes public transit; taxis and limos; bikesharing; carsharing (round-trip, one-way, and peer-to-peer); ridesharing (i.e., non-commerical services like carpooling and

the University of Washington Transportation Data Collaborative; it took place on March 12-14, 2018 in downtown Chicago, Illinois at the Radisson Blu Aqua Hotel, a space bedazzled with water drop-shaped sofas and lights. The event was billed as "an in-depth focus on public-private partnership development and relationship building across the sectors, including a reception, multiple networking breaks, and hands-on workshops to facilitate conversations." (Shared Use Mobility Summit Website). The recent growth in the market of "app-based" transportation options was palpable; the conference was sold out with 620 attendees from not only federal, state, and local transportation agency leadership, but also sales personnel from a range of bikeshare, scooter share, car share, and rideshare companies. Indeed, the event description also explicitly called out the conference as a market-making opportunity: "[The Summit features] opportunities for deal-making and partnership development with over 500 attendees…include[ing] city officials, commissioners from departments of transportation, transit agency leaders, and CEOs of shared mobility operators, tech companies, insurance providers, and leading automakers" (Shared Use Mobility Center 2018).

The workshop the day before the conference was a meeting of a two-year partnership between the Shared Use Mobility Center and the Federal Transportation Administration, called the Innovation and Knowledge Accelerator. This effort set out to convene grantee programs of the Mobility on Demand Sandbox project for real-time learning and reflection. Other grantee programs represented a broad swath of ideas, such as an effort to connect their transit cards to shared bikes in Chicago, an app integrating payments for public transit, bikeshare, and rideshare in Dallas, or an open-source integrated trip planner in Oregon.

The Innovation and Knowledge Accelerator Workshop was held in a series of conference rooms on March 11, the day before the conference was set to begin. Organizers broke out participants into small groups; I joined a session on data sharing. Grantee agencies reported similar problems as the agencies were experiencing in Seattle and Los Angeles; for instance a representative from the Dallas Area Rapid Transit reported in his presentation, "Negotiating acceptable agreements with the [rideshare companies] for shared ride service and access to data to evaluation the program has proven more difficult than anticipated." A representative from the

vanpooling); ridesourcing or ride-hailing; ride-splitting; scooter sharing (now often grouped with bikesharing under the heading of "micromobility"); shuttle services and "microtransit"; jitneys and dollar vans; and more" (Shared Use Mobility Center n.d.).

Oregon Tri-County Metropolitan Transportation District said that [rideshare] data agreements were their primary challenge. Many in the group reported that they had dealt with data sharing problems to date by asking for only a small amount of data or aggregate data assets from partners (Isaac 2017).

Figure 10. Chicago Radisson Blu Aqua Hotel, site of the Shared Use Mobility Summit

Against this backdrop of general challenges in gaining access to firm data as part of their ongoing initiatives and oversight, the conference also featured presentations from firms who were selling data to agencies as products. conference featured presentations from companies who create data as a by-product of their main business models---such as Allstate insurance company, whose subsidiary Arity sells data to governments originally generated from the in-car telematics drivers use to lower their insurance rates. After Arity's presentation, some questions from public employees in the audience signaled their uneasiness with the poor quality of this repurposed data against the needs that agencies usually made of it. As we packed up to leave the event, I talked with a woman my age who worked at a consultancy that conducts rigorous surveys of commuters to understand their behavior in a granular way—she said that firms usually present their data as the answer, but not to the problems that agencies most want to solve. Just as in conflict over the data that Lyft would be willing to share with King County and LA Metro, the questions in the audience indicated gaps between agencies' expectations for data, and the forms and formats in which companies were willing to provide it on their own terms. After years of an uneasy

relationship between rideshare companies and public transportation authorities, the conference represented a shift toward further collaboration.[55]

5.7 Conclusion

In this chapter, I have explored what is at stake in data sharing relationships with third party partners. Data sharing needs arise in part because agencies maintain their commitments to public interest goals, like accountability, transparency, and equity in access as defined by race as well as paratransit. At the heart of this disagreement were different visions for who owns data generated in the course of the partnership; Lyft asserted that data sharing was unworkable because of the data's privacy and competitiveness concerns, and both King County Metro and Los Angeles Metro asserted ownership claims to data, given the public investment in the program and the need to provide accountability over how public funds are spent. The agency requiring data from a new program ultimately had to turn to a partner that was willing to share data; this is an example of ways that access to "proprietary" data acquires increasing salience to public agencies, even in cases where they are procuring services via partnership instead of a vendor model.

This case also illustrates that data ownership claims do not determine who is ultimately able to access or control data, and that claims to data ownership can be overridden by other forms of compelled data disclosure. Data ownership is deployed as a means to both gain and bar access to data, and that multiple parties can assert ownership to the same data under different rationales. In spite of a set of continuing appeals that the data belonged to rideshare companies and should be protected as its intellectual property, the 2018 Washington Supreme Court decision that the City of Seattle had to disclose data collected by Lyft and Uber in response to a public records request shaped Lyft's unwillingness to share data for the Mobility on Demand Sandbox project. In this sense, public records laws impeded agency access to data, in that any industry information shared with regulators may be subject to requests that could unduly expose customer data to the public. This case illustrates the ways that data ownership is asserted both by public agencies like King County Metro and LA Metro, to get access to data, as well as by firms like Lyft and Uber to bar

[55] In a series of remarks the Executive Director of the Shared Use Mobility Center published after the Summit, one major takeaway of the conference was summarized as tracing a major shift from an adversarial role between agencies and rideshare firms to a collaborative one: “Cities are moving into the role of mobility brokers and convene in “mobility” (Feigon 2018).

others from obtaining it. In multiple vignettes in this chapter—from enabling Mobility on Demand, to oversight functions performed by the City of Seattle Department of Transportation, and the New York City Taxi and Limousine Commission—data sharing and access play a key role in the public sector's efforts to oversee the function of private firms under their jurisdiction.

Finally, this chapter illustrates that even as the "mobility" paradigm evinces an important shift from an *adversarial* relationship (in which public agencies attempt to enforce compliance with local laws) to *partnership* (in which public agencies work together with rideshare firms to enact local programs), data sharing came to play an important role in how accountability over rideshare firms was enforced. Through the lens of boundary formation drawn from institutional sociology (Mayrl and Quinn 2016), data sharing becomes a site of contestation over the boundaries of the state. Whether the state is regulating firms or partnering with them in a neoliberal sense, data sharing represents a form of enmeshment between the two that the firms attempted to prevent. Note that there are different ways to read the shift towards partnership between transportation agencies and firms: while in one sense it can be understood as a form of acceptance and even state sanction of the operation of the once-rogue rideshare industry, under another view the risks of further regulation invited by close data sharing posed a greater risk to participating firms than strategic benefit. Lyft's decision not to participate in the partnership due in part to its data sharing concerns underlines these continuing tensions. Importantly, this reading of data sharing as a practice of state re-formation indicates one sense in which the sphere of state influence is changing shape rather than diminishing, in contrast to readings of the neoliberal state as ceding ground to private firms.

Chapter 6. CONCLUSION

In this chapter, I reflect on contested claims for data sharing and access in government contracts, with a focus data ownership as a common rationale advanced my multiple parties when negotiating data control. I set out here to characterize how data ownership is enacted in practice, that is, made meaningful in its situated context. A grounded perspective on data ownership (and how it relates to access and control rights) can be instructive in this moment, when many advocates and policy proposals are calling for ownership-based regimes to data protection, remuneration, and control. Here I compare the meanings and effects (or lack thereof) that the term "ownership" takes on in the context of each case study and reflect on the implications of these findings for broader policy debates. I find that data ownership is intended on the ground to signify control over who is allowed to access data, and in this regard is associated with good data governance and rights protection. However, these claims are contested when multiple parties assert ownership to the same data; the promise of ownership as a means to control data is not borne out in practice. Indeed, my findings provide examples in which mediating factors like technological frictions prevented owners from accessing their own data or were "locked-in" to an existing information system that impeded their ability to make autonomous decisions about where to put data they owned. Finally, data ownership was not necessary to claim data access or control. I argue that my findings call into question the promise of contemporary policy and advocacy efforts to give individuals ownership over data about them, such as in recent state and federal bills (some of which have been adopted into law) or part of prominent political platforms.

6.1 PRIMARY FINDINGS

The previous case studies examine contested claims to data sharing and access in two case studies. Here, I revisit my research questions and answer them in each case.

6.1.1 *Findings from the ORCA-Vix vendor relationship*

The first case study reports on a fare card system that allows for easier transfer across different local public transit providers using single transit fare card called "one regional card for all," or "ORCA," and its eponymous inter-agency partnership in the Puget Sound region. In response to my first research question, "*What information systems and services are contractors asked to provide?*" I find that the agencies had contracted with a vendor called Vix in 2003 to be responsible for the sensors, data warehousing, payment systems, and customer service software needed to operate the system. The original vendor agreement specified that while the hardware and software belonged to the vendor, the data generated by fare transactions belonged to the agencies. The agreement also specified that vendor would be able to supply a copy of the agencies' data at any time. In response to my second research question, "*What roles do data sharing and access play in public sector contracting*?" I find that data sharing and access met multiple purposes; routine customer access needs were met by data that was readily available to the public agencies. Over time, the ORCA regional partnership also became interested in exporting its longitudinal data for use in long-term transportation planning and research. The chapter also explores claims on data access by institutional customers (e.g. local employers like Amazon and Microsoft) for the purpose of cost reduction, fraud prevention, and planning. It also describes the work of a local tech activist, Tyler Traver, to request access to code behind Vix's software in order to establish a public interest in inspecting software. In response to my third research question "*What technical, legal, and policy factors mediate data access and sharing in public sector contracting for information systems and services?*" I find that in this case, the practical and technical constraints on ORCA data access were the primary barrier to the partnering agencies' longitudinal data access, which required a $285,000 payment to the vendor and amendment to the original agreement. Institutional customers' data access was constrained by advocates' assertion that individual transportation data is sensitive and private to individual users, which found expression in amendments to the Public Records Act preventing employers from gaining access to all employees' detailed transportation records. The request for vendor software by Traver was denied on trade secret grounds, which the court upheld. In each case, the rationales used to deny or condition access to data and software were effective but varied greatly between specific requests.

6.1.2 *Findings from the King-County Metro-Lyft would-be partnership*

The second case study reports on data access and sharing in the public-private partnership behind the Mobility on Demand Sandbox First Mile Last Mile program, which was part of a joint partnership between King County Metro and Los Angeles Metro. In 2016, the Federal Transit Administration provided pilot funds for innovative programs between transit agencies and rideshare companies, and funded a joint proposal submitted by King County Metro (and its local partners), Los Angeles Metro, and Lyft for an ambitious pilot program to bring riders to and from their homes to connect them to transit hubs. The program stalled out after receiving federal funds before it could launch, due in part to Lyft's concerns about data sharing. In response to my first research question, "*What information systems and services are contractors asked to provide?*" I find that to enable the program and provide for its ongoing evaluation, Lyft was asked to both collect and provide detailed data about riders, routes, trips, and prices charged to agencies as part of its work to operate the first-mile last-mile service to hub stops. In response to my second research question, "*What roles do data sharing and access play in public sector contracting*?" I find that robust data sharing and access played multiple intersecting roles, including enabling agencies to assess the correct subsidy level, allowing researchers to evaluate the program and best practices, overseeing that drivers for subsidized trips were selecting the best routes, and ensuring that riders were using subsidized trips as part of a larger transit-enabled commute. In each respect, data sharing between Lyft and the public agencies was to play an essential role in the program.

Notably, A public-private partnership with rideshare companies would have been part of a larger turning point in the relationship between rideshare companies and local governments, but Lyft's reluctance to move forward with the Mobility on Demand Sandbox pilot was attributed by those close to the case as the company's concern over regulatory exposure. In the backdrop of this case, an ongoing public records request to the City of Seattle for data it held about Uber and Lyft was contested by both firms on the basis of trade secret. In response to my third research question, "*What technical, legal, and policy factors mediate data access and sharing in public sector contracting for information systems and services?*" I find that intellectual property assertions became a primary factor mediating access, which found expression in the legal arguments made by Uber and Lyft when they tried to prevent disclosure of data in an unrelated, contemporaneous court case in the Washington State Supreme Court. In response to my fourth research question "*Under what rationales do different actors assert or deny data access and sharing requests?*" I

find that both Lyft and King County Metro disagreed about which of the two entities would "own" data generated under the partnership, and each asserted data ownership as part of the rationale behind why data should be shared or withheld. Lyft's rationale for withholding data led to the dissolution of the proposed partnership. Ultimately the Washington Supreme Court decided that trade secret claims did not supersede legitimate public interests in the data, including research data access to assess equity in rideshare service provision.

6.2 Examining Data Ownership

In this chapter, I explore lessons learned across both cases with a focus on data ownership, given that it emerged as a key rationale for asserting and denying access to data. Rather than working with an a priori theory of what data ownership is, I consider how data ownership is enacted in practice by stakeholders on the ground and the arguments they use to support their ownership claims. By "enactment," I invoke ethnographic researcher Annemarie Mol rather than defining a concept seek to explore its multiplicity in situated use. Mol (2002) advances an approach to studying a seemingly stable concept, atherosclerosis, to show how it is brought into being and understood by different situated actors, such as doctors, patients, technicians, revealing the concept in its multiplicity. In this respect, rather than advancing my own definition of data ownership, I am interested in the ways that it operates on the ground, its contradictions, and how it is deployed by various stakeholders.

Previous work affirms data ownership as legally and conceptually fraught. Recall that in my literature review chapter, I reviewed data ownership literature as one form of private information, one that elegantly joins together different semiotic dimensions of the concept of "private" as both secrecy and property. I also traced ambiguity and contestation within the public/private dichotomy, finding that the conceptual challenges present in the dichotomy bear themselves out in interstitial spaces between the public sector and private sector, as well as public information and private information. As I traced information circulating in cross-sector contracting relationships, my empirical findings resonate with this work, particularly with Gal's (2002) assertion that the terms public and private are always asserting normative claims about the world. Understood as a normative claim, the following findings on how data ownership fares as a strategy for controlling access to data. I find that data ownership is not dispositive of outcomes and raise

questions about its increasing role in contemporary policy discussions about data rights and governance.

6.3 How Data Ownership Works in Context

In this section I summarize four findings about how data ownership works in practice and argue that it does not determine outcomes in practice with respect to who can access and control data. I focus my remarks here on data ownership in part because it is a rationale that emerged in multiple forms as I explored contestation in cross-sector data sharing and access, and in part because of its growing salience in policy discussions.

6.3.1 *Data ownership is intended to control access*

In both case studies, organizations sought to advance their own interests under the rhetoric of ownership, which I use here to encompass both explicit use of terms like "own" to describe an entity's relationship to data, as well as trade secret claims over data. Recall for example the way that Vix and the ORCA agencies outlined which aspects of the information system (hardware, software, data, user manuals, documentation) would belong to which entity. In this sense, data ownership is understood to be a way of asserting control in the form of rights to access, withhold, or share data. When data from firms in both case studies was requested in different public records requests, the affordances of the Washington Public Records Act (that is, to bar access to data on trade secret grounds) inclined firms in both cases towards asserting intellectual property claims to data as a means of withholding access. Notably, "owning" data is defined on the ground in a way that is distinct from "having" data, and the two are not coextensive; recall that the ORCA agencies "owned" longitudinal data generated by ORCA cards, but did not have direct access to it until after 2017. So, too, did the ORCA agencies possess a copy of the binary code requested by tech activist Tyler Traver, but were not allowed to grant access rights to it on the understanding that it was owned by Vix. Similar starting assumptions are evident in the case of Lyft and King County Metro, even as they were resolved differently as I will describe below.

6.3.2 *Data ownership performs good governance*

Writing data ownership into contracts lent the valence of certainty and control to data rights management; signed agreements appeared to fix data rights and good governance in place. In time, the language of these agreements may not provide the protections they intended. As the ORCA Next Generation System Architect described to me, data ownership was an early standard in whether a public agency was looking out for its interests in information systems:

> *"It's at a pivotal change in perception right now…. Owning the data used to be important. Now having the data and trying to do something with it is seen as important. They still don't know what, they still don't know how… the whole data governance thing is still raw and green right now. A lot of the decision-making [had] come from a point of minimum rights protection. So: 'We own the data.' 'Oh good.' The end." (interview, March 28 2018).*

Here, the respondent traces a maturation in thinking at the agencies as to how ownership relates to "doing something with it." He states that whereas before, ownership was widely considered to be good governance, with time that perception shifted to a sense it was more performative than actual. Comparing the two case studies highlights interesting similarities in this regard; King County Metro and LA Metro both had been adamant that they own data resulting from the pilot program, even to the point that they had to go find a new private sector partner willing to allow this. A key difference however is that King County Metro and LA Metro asserted that they would own data in the sense of also "having" it—direct access to data was another key tenet of their negotiating position (one which disinclined them to work with our university-based intermediary).

The primary sense in which data ownership *performs* data access and control rights is that these rights are not made actual. In the case of the ORCA regional partnership, the agencies "owned" data that they could not access without considerable additional expense; practical and technological constraints on data management trumped ownership. In a larger sense, this example also points to the way that the imagined virtues of an ownership paradigm (of autonomy, sovereignty, and choice) is overridden in practice by real frictions like technological lock-in. Respondents within and close to the ORCA agencies were clear that avoiding lock-in was a primary motivator behind opening the contracting process for a "Next Generation" modular, interoperable system to replace the current single-vendor paradigm. Ownership is also performative in the multiple examples by which entities asserted ownership over data but were not able to assume meaningful control.

6.3.3 *Owners may not be able to prevent others access*

Where an entity has asserted ownership, it may not be able to reliably bar others' accessing data. This can occur in several ways. First, multiple parties can assert ownership over the same data. When negotiations failed between King County Metro and Lyft, neither party could agree on who owned data, and used competing rationales to support their conflicting arguments. King County Metro, for its part, asserted that the subsidy they would pay partners for the service enabled the data to exist and should grant them control; as the Innovation Mobility Program Manager explained, "To me if we are paying for the rides, then we own that data as much as anyone else" (interview, April 16 2018, emphasis added). Note the use of the phrase "as much as anyone else," which leaves room for the possibility that data can be jointly owned by different entities. While Lyft is not on the record with their rationale for data ownership in the Mobility on Demand Project, they asserted their data to be trade secret in a related court case they had brought against a public records requestor and the City of Seattle at the same time. Second, other interests can trump proprietary interests for data access in compelled disclosure. In the second case study, despite multiple rounds of appeals, the Washington Supreme Court ruled that even as it recognized that both rideshare companies' assertion that requested records could be trade secret, they could nevertheless be disclosed as public records given the public interest in research data access, accountability, and equity. Altogether, ownership as a rationale was not a reliable data protection method.

6.3.4 *Data ownership is not necessary for access, control, or sharing*

Finally--and especially interesting for our purposes at the time in light of our work building a data sharing intermediary—data ownership claims were neither necessary nor sufficient to make effective data sharing and access claims. The aforementioned case between Uber, Lyft, and the City of Seattle recognized a public interest in certain records irrespective of their ownership. More routinely, the myriad data access requests from ORCA card institutional customers wanting to examine their budgets, card utilization rates, and misuse are prime examples of the kinds of data sharing requests that both public agencies and private firms navigate in the absence of an assertion

of ownership seeking to bolster those claims. Viewed from another perspective, data access claims tend to rely on a common set of themes that include access expected in exchange for paying for some part of the system, access to serve the public interest (such as equity in service provision), and access to ensure parties' proper oversight and accountability. Taken together, the original impetus of the University of Washington Transportation Data Collaborative was to be able to address researchers' and agencies' stated needs for data without providing them with direct access to the data itself. Our proposed data sharing intermediary did not obviate the need to navigate ownership claims to data (particularly given that in the Washington state legal context, the exemption used for intellectual property is a reliable means to protect data from disclosure). However, it did help to uncouple the competing interests advanced in data sharing disputes from the assertion of data ownership as a rationale for enabling or constraining that access. For this reason, I have come to view data ownership as hollow, and its stated aims as addressable by more granular discussions of how data access can be made available and under what mutually agreeable terms.[56]

6.4 Policy Implications: Why It Matters

My empirical cases suggest that "ownership" may not ensure that data is governed as owners (and subjects) intend, nor is it determinative of the rights and obligations that the metaphor suggests. I turn briefly here to examine data ownership's contemporary use as a means of advancing digital rights. A close analysis of how data ownership works in practice is important in light of its increasing salience as a policy proposal. Indeed, data ownership is considered to be an important path forward for protecting consumers' digital rights. In their book ``The End of Ownership: Personal Property in the Digital Economy'' Perzanowski and Shultz (2017) argue in favor of private ownership of digital goods, given the dilution of the term in contemporary end user licensing agreements and terms of service. In 2019 the U.S. Senate proposed a bill that ``allows users to access the personal data that social media companies have compiled, and it transfers data property

[56] I owe this perspective to Jan Whittington, whose expertise in transaction cost economics deeply informed the impetus behind the University of Washington Transportation Data Collaborative; this conceptual heritage takes contracts as a starting point to account for and divert harms.

rights back into the hands of the users" called the Own Your Own Data Act (cite S. 806). [57] In the same year, both Oregon and Maryland passed bills (drafted by blockchain-powered monetization company Hu-manity.co) that would allow patients to own their healthcare data and receive payments from its sale (Beebe 2019); notably the American Civil Liberties Union (ACLU) describes these state bills as a ``trojan horse" for new industry actors to profit from a role as data broker intermediaries (Marlow 2019). In 2019, Democratic presidential primary contender Andrew Yang also advocated for individual property rights over data.

Beyond my field sites, I understand these findings on the limits of data ownership as useful for interrogating claims by large technology platforms like Facebook and 23andMe that users "own" their data. In response to rising concerns about data governance, large technology firms have assured the public that users own their own data. In April 2018 a U.S. Senator asked Mark Zuckerberg, CEO of Facebook, "You consider my personally identifiable data the company's data, not my data. Is that it?" To which Zuckerberg responded, "No, senator. Actually, the first line of our Terms of Service say that *you control and own* the information and content that you put on Facebook" (The Washington Post 2018). Other firms have also asserted that users own their own data. In October that year, the Global Privacy Officer of genetics company 23andMe, Kate Black, stated in an interview, ``You always own your data. You do have to give us a license in order to process it, analyze your sample and provide you with reports back… but you do retain entire ownership rights over that" (Glaser and Oremus 2018). In both cases, data ownership is advanced as a remedy to user concerns about data.

Even when users are said to own personal data, both firms provide important clarifications as to how data is retained, accessed, and shared beyond users' control. In the same Senate hearing, Zuckerberg describes the granular control that users have over how visible their posts are to other users—--even as he indicates that there is no way to opt out of data sharing with advertisers and law enforcement. Similarly, 23andMe's Black outlines that users have control over whether and how their information shared with research or other entities. She also clarifies that when users' samples are analyzed they grant an ``irrevocable" right to the company ("just like every other

[57] "This bill requires a social media company to (1) provide a user with the ability to obtain and export their user data, along with any analysis of such data by the company; and (2) ensure that each user knowingly and willingly licenses their data to the company during the registration of an account." Summary of S. 806 Own your Own Data Act by the Congressional Research Service.

laboratory testing site") to retain ``a small subset of your information in order to maintain our clinical quality standards and regulatory requirements" (Glaser and Oremus 2018). These examples illustrate that data owners are not solely responsible for controlling it and affirm the findings of my empirical work on its flaws as an approach to data control.

6.5 Conclusion

This chapter recounted lessons learned across both cases with a focus on data ownership as it emerged as a key rationale for asserting and denying access to data. Rather than working with an a priori theory of what data ownership is, I considered how data ownership is enacted in practice by stakeholders on the ground and the arguments they use to support their ownership claims. I found that data ownership is used as a tactic to try to control who can access and share data. The agencies that assert ownership claims to the data they use are trying to protect their interests and rights; from a rights protection perspective, data ownership is associated with good governance. Indeed, ownership assertions can be multiple and contested; public agencies and the private firms both claimed ownership to the same data as a tactic to elicit or deny data access in both case studies, as well as indirectly by individual records requestors. I also find this positive valence to be performative—as data ownership alone did not ensure that the ORCA agencies were able to get a copy of their data on demand, but instead needed to pay additional overhead to navigate technical and practical constraints on data access. In that sense, data owners do not have the autonomy the metaphor implies to place their data where they like, but instead can be locked-in to the system where data is stored. These claims do not trump other interests in compelled disclosure for data, and neither were they necessary or sufficient for gaining access to data. Having explored the implications of these findings for current policy debates, I raise questions about the effectiveness of relying on ownership claims as part of a strategy to defend individuals' digital rights.

BIBLIOGRAPHY

ACLU Washington. 2010. "Travel Without Tracking: Employer Access to ORCA Info Restricted Travel Without Tracking: Employer Access to ORCA Info Restricted," January 7, 2010. https://www.aclu-wa.org/news/travel-without-tracking-employer-access-orca-info-restricted.

Aktinson, Paul, and M. Hammersley. 1998. "Ethnography and Participant Observation." *Strategies of Qualitative Inquiry. Thousand Oaks: Sage*, 248–261.

Andrejevic, Mark. 2014. "Big Data, Big Questions| The Big Data Divide." *International Journal of Communication* 8: 17.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks." *ProPublica*, May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Applebaum, Jacob. 2012. "Digital Anti-Repression & Mobile Security, Part One." Presented at the Occupy Wall Street, Walker Stage, April 26. https://www.youtube.com/.

Arendt, Hannah. 2013. *The Human Condition*. University of Chicago Press.

Ashton, Philip, Rachel Weber, and Matthew Zook. 2017. "The Cloud, the Crowd, and the City: How New Data Practices Reconfigure Urban Governance?" *Big Data & Society* 4 (1): 2053951717706718. https://doi.org/10.1177/2053951717706718.

Bailey, Joe. 2000. "Some Meanings Ofthe Private'in Sociological Thought." *Sociology* 34 (3): 381–401.

Barns, Sarah. 2016. "Mine Your Data: Open Data, Digital Strategies and Entrepreneurial Governance by Code." *Urban Geography* 0 (0): 1–18. https://doi.org/10.1080/02723638.2016.1139876.

Benjamin, Ruha. 2019. "Race after Technology: Abolitionist Tools for the New Jim Code." *Social Forces*.

Benn, Stanley I., and Gerald F. Gaus. 1983. *Public and Private in Social Life*. Taylor & Francis.

Beswick, Samuel. 2016. "The Decline of the Fish/Mammal Distinction." *University of Pennsylvania Law Review Online* 165: 91.

Bevir, Mark. 2011. "Governance and Governmentality after Neoliberalism." *Policy & Politics* 39 (4): 457–71. https://doi.org/10.1332/030557310X550141.

Blomley, Nicholas. 2005. "Flowers in the Bathtub: Boundary Crossings at the Public–Private Divide." *Geoforum* 36 (3): 281–296.

Bowen, Glenn A. 2009. "Document Analysis as a Qualitative Research Method." *Qualitative Research Journal* 9 (2): 27.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15 (5): 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Brenner, Neil, and Nik Theodore. 2002. "Cities and the Geographies of 'Actually Existing Neoliberalism.'" *Antipode* 34 (3): 349–79. https://doi.org/10.1111/1467-8330.00246.

Brock, Anke M., Jon E. Froehlich, João Guerreiro, Benjamin Tannert, Anat Caspi, Johannes Schöning, and Steve Landau. 2018. "Sig: Making Maps Accessible and Putting Accessibility in Maps." In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, SIG03. ACM.

Brown, J Pat. 2017. "The Sounds of Sound Transit." *Muckrock*, September 11, 2017. https://www.muckrock.com/news/archives/2017/sep/11/sounds-sound-transit/.

Brown, Trevor L., and Matthew Potoski. 2003. “Managing Contract Performance: A Transaction Costs Approach.” *Journal of Policy Analysis and Management* 22 (2): 275–297.
Bruckner, Matthew Adam. 2018. “The Promise and Perils of Algorithmic Lenders’ Use of Big Data.” *Chi.-Kent L. Rev.* 93: 3.
Buchanan, James M. 1965. “An Economic Theory of Clubs.” *Economica* 32 (125): 1–14.
Buolamwini, Joy, and Timnit Gebru. 2018. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” In *Proceedings of Machine Learning Research*, 81:15. New York, NY.
Burgess, Michael M. 2014. “From ‘Trust Us’ to Participatory Governance: Deliberative Publics and Science Policy.” *Public Understanding of Science* 23 (1): 48–52. https://doi.org/10.1177/0963662512472160.
Burki, Shahid Javed, Guillermo Perry, and William R. Dillinger. 1999. *Beyond the Center: Decentralizing the State*. World Bank Publications.
Calo, Ryan, and Alex Rosenblat. 2017. “The Taking Economy: Uber, Information, and Power Essay.” *Columbia Law Review* 117 (6): [i]-1690.
Carey, Christopher. 2020. “How Los Angeles Took Control of Its Mobility Data.” *Cities Today*, July 1, 2020. https://cities-today.com/how-los-angeles-took-control-of-its-mobility-data/.
Casey, Bryan. 2017. “Uber’s Dilemma: How the ADA May End the On-Demand Economy.” *U. Mass. L. Rev.* 12: 124.
Castells, Manuel. 2013. *Communication Power*. OUP Oxford.
Clemens, Elisabeth S., and Doug Guthrie. 2010. *Politics and Partnerships: The Role of Voluntary Associations in America’s Political Past and Present*. University of Chicago Press.
Coase, Ronald H. 1937. “The Nature of the Firm.” *Economica* 4 (16): 386–405.
Cohen, Jean L. 1997. “Rethinking Privacy: Autonomy, Identity, and the Abortion Controversy.” *Public and Private in Thought and Practice: Perspectives on a Grand Dichotomy*, 133–165.
Collins, Flannary. 2017. “New PRA Legislation: To Boldly Go Where the PRA Hasn’t Gone Before.” *Municipal Research and Services Center*, May 30, 2017. http://mrsc.org/Home/Stay-Informed/MRSC-Insight/May/New-PRA-Legislation.aspx.
Conti-Cook, Cynthia H. 2017. “Open Data Policing.” *Geo. LJ Online* 106: 1.
Crump, Catherine. 2016. “Surveillance Policy Making by Procurement.” *Wash. L. Rev.* 91: 1595.
Delaney, David. 2010. *The Spatial, the Legal and the Pragmatics of World-Making: Nomospheric Investigations*. Routledge.
Dijck, José van. 2013. *The Culture of Connectivity: A Critical History of Social Media*. Oxford ; New York: Oxford University Press.
Drummond, Lisa BW. 2000. “Street Scenes: Practices of Public and Private Space in Urban Vietnam.” *Urban Studies* 37 (12): 2377–2391.
Dunleavy, Patrick, Helen Margetts, Simon Bastow, and Jane Tinkler. 2006. “New Public Management Is Dead—Long Live Digital-Era Governance.” *Journal of Public Administration Research and Theory* 16 (3): 467–494.
Edwards, Paul N. 2003. “Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems.” *Modernity and Technology* 1.
Ellickson, Robert. 1982. “Cities and Homeowners Associations.” *University of Pennsylvania Law Review* 130 (6): 1519.
Elshtain, Jean Bethke. 1981. *Public Man, Private Woman*. JSTOR.

Emerson, Thomas I. 1976. “Legal Foundations of the Right to Know.” *Wash. ULQ*, 1.
EPIC. n.d. “State FOIA: Secret DNA Forensic Source Code.” Electronic Privacy Information Center. https://epic.org/state-policy/foia/dna-software/.
Erhardt, Gregory D., Sneha Roy, Drew Cooper, Bhargava Sana, Mei Chen, and Joe Castiglione. 2019. “Do Transportation Network Companies Decrease or Increase Congestion?” *Science Advances* 5 (5): eaau2670.
Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin’s Press.
Fairfield, Paul. 2005. *Public/Private*. Rowman & Littlefield. https://books.google.com/books?hl=en&lr=&id=lyOwkbNw9EsC&oi=fnd&pg=PA1&dq=public/private+fairfield&ots=ytiWeZQGtZ&sig=rvNHjKGCtwRCFBgYzJm_fZ0kyg8.
Federal Transit Administration. 2019. “Mobility on Demand (MOD) Sandbox Program.” https://www.transit.dot.gov/research-innovation/mobility-demand-mod-sandbox-program.
Feeney, Matthew. 2015. “Is Ridesharing Safe?” *Cato Institute Policy Analysis*, no. 767.
Feigon, Sharon. 2018. “Some Thoughts on the Evolution of Shared Mobility from the 2018 National Shared Mobility Summit.” Shared Use Mobility Center. https://sharedusemobilitycenter.org/thoughts-evolution-shared-mobility-2018-national-shared-mobility-summit/.
Ferris, Brian, Kari Watkins, and Alan Borning. 2010. “OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1807–1816. ACM.
Foucault, Michel. 1978. *The History of Sexuality*. 1st American ed. New York: Pantheon Books.
Fraser, Nancy. 1998. “From Redistribution to Recognition? Dilemmas of Justice in a ‘Post-Socialist’ Age.” *Feminism and Politics*, 430–460.
Freeman, Jody. 2000. “Private Role in the Public Governance, The.” *New York University Law Review* 75: 543.
Friedman, Milton. 1962. “Capitalism and Freedom.” *University of Chicago*.
Friedman, Milton, and Rose Friedman. 1990. *Free to Choose: A Personal Statement*. Houghton Mifflin Harcourt.
Gal, Susan. 2002. “A Semiotics of the Public/Private Distinction.” *Differences: A Journal of Feminist Cultural Studies* 13 (1): 77–95.
Gavison, Ruth. 1992. “Feminism and the Public/Private Distinction.” *Stanford Law Review*, 1–45.
Girth, Amanda M., Amir Hefetz, Jocelyn M. Johnston, and Mildred E. Warner. 2012. “Outsourcing Public Service Delivery: Management Responses in Noncompetitive Markets.” *Public Administration Review* 72 (6): 887–900.
Glaser, April, and Will Oremus. 2018. *If Then Podcast: Who Owns Your DNA Data?* If Then Podcast. https://www.stitcher.com/podcast/slate-magazine-daily-podcast/e/56896326.
Gosseries, Axel, and Tom Parr. 2017. “Publicity.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2017. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2017/entries/publicity/.
Graham, Stephen, and Simon Marvin. 2001. *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*. Psychology Press. https://books.google.com/books?hl=en&lr=&id=JA9dNAR5evEC&oi=fnd&pg=PA1&dq=splintering+urbanism&ots=f3O083Xskp&sig=cvD1m6fanNo4o8xyC1tOS0CWuzk.

Griffin, Greg P., Gretchen Stoeltje, Nathan Jones, Nicholas S. Wood, and Chris Simek. 2016. "Recruiting Carpoolers: Dynamic Ridesharing with Incentives in Central Texas."
Grossman, Alice, and Paul Lewis. 2020. "Data on Demand: A Case Study in the Los Angeles and Puget Sound Regions," February. https://trid.trb.org/view/1685039.
Gutman, David. 2018. "Uber and Lyft May Have to Disclose Seattle Data They Claim Secret, Supreme Court Rules." *The Seattle Times*, May 31, 2018. https://www.seattletimes.com/seattle-news/transportation/uber-and-lyft-may-have-to-disclose-data-they-claim-secret-supreme-court-rules/.
Hallenbeck, Mark E., Eric Howard, Dmitri Zyuzin, Ryan P. Avery, and Miki Verma. 2017. "Use of Electronic Fare Transaction Data for Corridor Planning." Washington (State). Dept. of Transportation. Research Office.
Hartzog, Woodrow. 2019. "The Public Information Fallacy." *BUL Rev.* 99: 459.
Harvey, David. 2005. *A Brief History of Neoliberalism*. OUP Oxford.
Hawkins. 2017. "Uber Sues Seattle over Law Allowing Drivers to Unionize." *The Verge*, 2017. https://www.theverge.com/2017/2/6/14524792/uber-lawsuit-seattle-law-drivers-unionize.
———. 2019. "Uber Launches a New Tool to Help Cities Track Average Vehicle Speeds by Street," May 14, 2019. https://www.theverge.com/2019/5/14/18623149/uber-movement-average-car-speed-data-cities.
Hayek, F. A. 1945. "The Use of Knowledge in Society." *The American Economic Review* 35 (4): 519–30.
Henao, Alejandro, and Wesley E. Marshall. 2019. "The Impact of Ride-Hailing on Vehicle Miles Traveled." *Transportation* 46 (6): 2173–2194.
Hettinger, Edwin C. 1989. "Justifying Intellectual Property." *Philosophy & Public Affairs*, 31–52.
Hirschman, Albert O. 1982. *Shifting Involvements: Private Interest and Public Action*. Princeton University Press.
Hood, Christopher. 1991. "A Public Management for All Seasons?" *Public Administration* 69 (1): 3–19.
———. 1995. "The 'New Public Management' in the 1980s: Variations on a Theme." *Accounting, Organizations and Society* 20 (2–3): 93–109.
Isaac, Mike. 2017a. "How Uber Deceives the Authorities Worldwide." *The New York Times* 3: 17.
———. 2017b. "Uber Extends an Olive Branch to Local Governments: Its Data." *The New York Times*, January 8, 2017. https://www.nytimes.com/2017/01/08/technology/uber-movement-traffic-data.html.
Isaac, Mike, Nick Wingfield, and Noam Scheiber. 2015. "Seattle Considers Measure to Let Uber and Lyft Drivers Unionize." *The New York Times* 13.
Jorge, Diana, and Gonçalo Correia. 2013. "Carsharing Systems Demand Estimation and Defined Operations: A Literature Review." *European Journal of Transport and Infrastructure Research* 13 (3).
Kamargianni, Maria, Weibo Li, Melinda Matyas, and Andreas Schäfer. 2016. "A Critical Review of New Mobility Services for Urban Transport." *Transportation Research Procedia* 14: 3294–3303.
Kant, Immanuel. 1915. *Perpetual Peace*.
Kennedy, Duncan. 1981. "Stages of the Decline of the Public/Private Distinction." *U. Pa. L. Rev.* 130: 1349.

Keynes, John Maynard. 2016. *General Theory of Employment, Interest and Money*. Atlantic Publishers & Dist.
King County Metro. 2017. “Metro and Via to Join Forces on Shared Rides to and from Select Transit Stations.” 2017. https://www.metro.net/news/simple_pr/metro-and-join-forces-shared-rides-and-select-tran/.
Kitchin, Rob. 2014. “The Real-Time City? Big Data and Smart Urbanism.” *GeoJournal* 79 (1): 1–14.
Klare, Karl E. 1982. “The Public/Private Distinction in Labor Law.” *University of Pennsylvania Law Review* 130 (6): 1358–1422.
Klein, Lawrence Eliot. 1995. “Gender and the Public/Private Distinction in the Eighteenth Century: Some Questions about Evidence and Analytic Procedure.” *Eighteenth-Century Studies* 29 (1): 97–109. https://doi.org/10.1353/ecs.1995.0010.
Koppenjan, Joop FM, and Bert Enserink. 2009. “Public–Private Partnerships in Urban Infrastructures: Reconciling Private Sector Participation and Sustainability.” *Public Administration Review* 69 (2): 284–296.
Larkin, Brian. 2013. “The Politics and Poetics of Infrastructure.” *Annual Review of Anthropology* 42: 327–343.
Lawler, Ryan. 2014. “Seattle Deals A Blow To Uber And Lyft By Limiting The Number Of Ride-Sharing Drivers On The Road.” *TechCrunch*, March 17, 2014. https://techcrunch.com/2014/03/17/seattle-hates-ride-sharing/.
Lessig, Lawrence. 1998. “The New Chicago School.” *The Journal of Legal Studies* 27 (S2): 661–91. https://doi.org/10.1086/468039.
———. 2009. *Code: And Other Laws of Cyberspace*. ReadHowYouWant. com.
Lewis, George Cornewall, and Roland Knyvet Wilson. 1877. *Remarks on the Use and Abuse of Some Political Terms*. J. Thornton.
Li, Sen, Hamidreza Tavafoghi, Kameshwar Poolla, and Pravin Varaiya. 2019. “Regulating TNCs: Should Uber and Lyft Set Their Own Rules?” *Transportation Research Part B: Methodological* 129: 193–225.
Lin, Jane, Sandeep Sasidharan, Shuo Ma, and Ouri Wolfson. 2016. “A Model of Multimodal Ridesharing and Its Analysis.” In *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, 1:164–173. IEEE.
Linder, Stephen H. 1999. “Coming to Terms with the Public-Private Partnership a Grammar of Multiple Meanings.” *American Behavioral Scientist* 43 (1): 35–51.
Ma, Shuo, Yu Zheng, and Ouri Wolfson. 2014. “Real-Time City-Scale Taxi Ridesharing.” *IEEE Transactions on Knowledge and Data Engineering* 27 (7): 1782–1795.
MacKinnon, Catharine A. 1989. *Toward a Feminist Theory of the State*. Harvard University Press.
Mapelli, Elizabeth A. 2017. “Inadequate Accessibility: Why Uber Should Be a Public Accommodation Under the Americans With Disabilities Act.” *Am. UL Rev.* 67: 1947.
Marshall, Aarian. 2019. “New York City Flexes Again, Extending Cap on Uber and Lyft. Wired.” *Wired*, 2019. https://www.wired.com/story/new-york-city-flexes-extending-cap-uber-lyft/.
Martin, Denise, Mary Jane Spink, Pedro Paulo Gomes Pereira, Denise Martin, Mary Jane Spink, and Pedro Paulo Gomes Pereira. 2018. “Multiple Bodies, Political Ontologies and the Logic of Care: An Interview with Annemarie Mol.” *Interface - Comunicação, Saúde, Educação* 22 (64): 295–305. https://doi.org/10.1590/1807-57622017.0171.

Marx, Gary T. 2001. “Murky Conceptual Waters: The Public and the Private.” *Ethics and Information Technology* 3 (3): 157–169.
Mayrl, Damon, and Sarah Quinn. 2016. “Defining the State from within: Boundaries, Schemas, and Associational Policymaking.” *Sociological Theory* 34 (1): 1–26. https://doi.org/10.1177/0735275116632557.
McNeill, Donald. 2015. “Global Firms and Smart Technologies: IBM and the Reduction of Cities.” *Transactions of the Institute of British Geographers* 40 (4): 562–74. https://doi.org/10.1111/tran.12098.
Mehta, Nayantara. 2016. “Ensuring Fairness in Background Checks for On-Demand Work.” *National Employment Law Project*.
Meiklejohn, Alexander. 1948. *Free Speech and Its Relation to Self-Government*. The Lawbook Exchange, Ltd.
Miletich, Steve. 2015. “Group Pushing for Public Records, More Police Accountability,” November 22, 2015. https://www.theolympian.com/news/politics-government/article37251510.html.
Mill, John Stuart. 1869. *On Liberty*. Longmans, Green, Reader, and Dyer.
Miller, Fred. 2017. “Aristotle’s Political Theory.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2017. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2017/entries/aristotle-politics/.
Miraftab, Faranak. 2004. “Public-Private Partnerships: The Trojan Horse of Neoliberal Development?” *Journal of Planning Education and Research* 24 (1): 89–101.
Mitchell, Timothy. 1991. “The Limits of the State: Beyond Statist Approaches and Their Critics.” *The American Political Science Review* 85 (1): 77. https://doi.org/10.2307/1962879.
Mnookin, Robert H. 1981. “Public/Private Dichotomy: Political Disagreement and Academic Repudiation.” *University of Pennsylvania Law Review* 130: 1429–40.
Mol, Annemarie. 2002. *The Body Multiple: Ontology in Medical Practice*. Duke University Press. https://books.google.com/books?hl=en&lr=&id=qdltz8N9qSEC&oi=fnd&pg=PR5&dq=body+multiple&ots=wHvEAPcSze&sig=TfY2t0YTGJ-m7nMDKKQBq3FXUwE.
Montjoye, Yves-Alexandre de, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. “Unique in the Crowd: The Privacy Bounds of Human Mobility.” *Nature Srep.* 3. http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+mediaredef+jason+hirschhorn%E2%80%99s+Media+ReDEFined.
Moon, Mariella. 2017. “Portland Finds Uber Used ‘Greyball’ to Evade 16 Authorities.” *Engadget*, September 15, 2017. https://www.engadget.com/2017-09-15-portland-probe-uber-greyball-results.html.
Morgan, Kimberly J., and Ann Shola Orloff. 2017. *The Many Hands of the State: Theorizing Political Authority and Social Control*. Cambridge University Press.
Morris, Stephen, and Hyun Song Shin. 2002. “Social Value of Public Information.” *American Economic Review* 92 (5): 1521–1534.
Murray, David. 2015. “Queensland Authorities Confirm That a Miscode Affects DNA Evidence in Criminal Cases.” *Courier Mail*, March 20, 2015. http://www.couriermail.com.au/news/queensland/queensland-authorities-confirm-miscode-affects-dna-evidence-in-criminal-cases/news-

story/833c580d3f1c59039efd1a2ef55af92b?=.
Nadleman, Jessica, and Erica Franklin. 2018. “Navigating Third Party Relationships Under the PRA.” Washington State Association of Municipal Attorneys: Seattle City Attorney’s Office. https://wsama.org/index.asp?SEC=7340814D-1D4F-42CA-A241-423A05486E77&DE=5812E30F-2481-4DB7-A1E0-02BB2A79A86B.
Newell, Bryce Clayton. 2013. “Local Law Enforcement Jumps on the Big Data Bandwagon: Automated License Plate Recognition Systems, Information Privacy, and Access to Government Information.” *Browser Download This Paper*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2341182.
Newman, Janet, and John Clarke. 2009. *Publics, Politics and Power: Remaking the Public in Public Services*. Sage.
Nissenbaum, Helen. 2004. “Privacy as Contextual Integrity.” *Washington Law Review* 79 (1).
———. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.
North, Douglass Cecil. 1990. *Institutions, Institutional Change and Economic Performance. Edn. Political Economy of Institutions and Decisions*. Cambridge University Press.
Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. “Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations.” *Science* 366 (6464): 447–453.
O’Brien, Sara, Neil Black, Curt Devine, and Drew Griffin. 2018. “CNN Investigation: 103 Uber Drivers Accused of Sexual Assault or Abuse.” *CNNTech, April* 30.
Okin, Susan Moller. 1989. *Justice, Gender, and the Family*. Basic books. https://books.google.ca/books?hl=en&lr=&id=uZvSDQAAQBAJ&oi=fnd&pg=PR1&dq=Justice,+Gender+and+the+Family&ots=7NNbuQrgjd&sig=J_xf8vJO6iguiTCOuaUUlocpINs.
Osborne, David, and Ted Gaebler. 1992. “Reinventing Government: How the Entrepreneurial Spirit Is Transforming Government.” *Reading Mass. Adison Wesley Public Comp*.
Oxford English Dictionary. n.d. “Private, Adj.1, Adv., and n.” In *OED Online*. Oxford University Press. Accessed February 20, 2018a. http://www.oed.com/view/Entry/151601.
———. n.d. “Public, Adj. and n.” In *OED Online*. Oxford University Press. Accessed August 4, 2017b. http://www.oed.com.offcampus.lib.washington.edu/view/Entry/154052.
Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.
Peck, Jamie, and Adam Tickell. 2002. “Neoliberalizing Space.” *Antipode* 34 (3): 380–404.
Pigou, Arthur Cecil. 2013. *The Economics of Welfare*. Palgrave Macmillan.
Pitkin, Hanna Fenichel. 1981. “Justice: On Relating Private and Public.” *Political Theory* 9 (3): 327–52.
Poitras, Laura. 2014. *Citizenfour*. Documentary. Radius-TWC. https://citizenfourfilm.com/.
Posner, Richard A. 1981. “The Economics of Privacy.” *The American Economic Review* 71 (2): 405–9.
Prainsack, Barbara. 2019. “Logged out: Ownership, Exclusion and Public Value in the Digital Data and Information Commons.” *Big Data & Society* 6 (1): 2053951719829773.
Randall, Ronald, James A. Woods, and Robert G. Martin. 2008. “Racial Representativeness of Juries: An Analysis of Source List and Administrative Effects on the Jury Pool.” *Justice System Journal* 29 (1): 71–84.
Robinson, Sean. 2015. “Group Sues Tacoma Police over Stingray Agreement.” *Tacoma News*

Tribune, October 2, 2015. http://www.thenewstribune.com/news/politics-government/article37341000.html.
Romany, Celina. 1993. "Women as Aliens: A Feminist Critique of the Public/Private Distinction in International Human Rights Law." *Harvard Human Rights Journal* 6: 87.
Ruud, Candice. 2017. "Tacoma Hit with $50,000 Fine plus Legal Fees over Stingray Nondisclosure Agreement." *The Tacoma News Tribune*, March 3, 2017. https://www.thenewstribune.com/news/politics-government/article136371963.html.
Sadowski, Jathan. 2016. "Selling Smartness Visions and Politics of the Smart City." ARIZONA STATE UNIVERSITY. http://gradworks.umi.com/10/14/10144453.html.
Samuelson, Pamela. 2000. "Privacy as Intellectual Property?" *Stanford Law Review*, 1125–1173.
Schwartz, Paul M. 2004. "Property, Privacy, and Personal Data." *Harvard Law Review* 117 (7): 2056–2128. https://doi.org/10.2307/4093335.
Sclar, Elliott. 2001. *You Don't Always Get What You Pay for: The Economics of Privatization*. Cornell University Press.
Sclar, Elliott D., K. H. Schaeffer, and Robert Brandwein. 1989. *The Emperor's New Clothes: Transit Privatization and Public Policy*.
Seattle Department of Transportation. 2017. "New Mobility Playbook." City of Seattle. https://newmobilityseattle.info/playbook.
Sennett, Richard. 2017. *The Fall of Public Man*. W. W. Norton & Company.
Shaheen, Susan, Adam Cohen, Adam Stocker, and Elliot Martin. 2019. "Mobility on Demand: A Smart, Sustainable, and Equitable Future." *Transportation Research Circular*, no. E-C244.
Shapiro, Carl, and Hal R. Varian. 1998. *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.
Shared Use Mobility Center. n.d. "What Is Shared Mobility?" https://sharedusemobilitycenter.org/what-is-shared-mobility/.
———. 2018. "2018 National Shared Mobility Summit." https://sharedusemobilitycenter.org/event/2018-national-shared-mobility-summit/.
Söderström, Ola, Till Paasche, and Francisco Klauser. 2014. "Smart Cities as Corporate Storytelling." *City* 18 (3): 307–20.
Solove, Daniel J. 2008. *Understanding Privacy*.
Squires, Gregory, M. Gottdiener, and C. Pickvance. 2012. "Partnership and the Pursuit of the Private City." *The Urban Sociology Reader*, 118–126.
Star, S. L. 1999. "The Ethnography of Infrastructure." *American Behavioral Scientist* 43 (3): 377–91. https://doi.org/10.1177/00027649921955326.
Stark, Andrew. 2010. *Drawing the Line: Public and Private in America*. Brookings Institution Press. https://books.google.com/books?hl=en&lr=&id=6xj5C9m_iU0C&oi=fnd&pg=PP1&dq=Drawing+the+line+:+public+and+private+in+America&ots=qJyJv6FZkn&sig=rMjVP2wnBMtQk1l6E1oIv9FfK6g.
Starr, Paul. 1990. "The New Life of the Liberal State." In *Public Enterprise and Privatization*. Boulder, CO: Westview Press.
Sunshine Committee. 2019. "2019 Annual Report." State of Washington Office of the Attorney General Public Records Exemptions Accountability Committee. https://agportal-s3bucket.s3.amazonaws.com/uploadedfiles/Another/About_the_Office/Open_Government/2019-Sunshine-Committee-Annual-Report.pdf.

Susarla, Anjana, Anitesh Barua, and Andrew B. Whinston. 2009. “A Transaction Cost Perspective of the ‘Software as a Service’ Business Model.” *Journal of Management Information Systems* 26 (2): 205–40. https://doi.org/10.2753/MIS0742-1222260209.
Tashea, Jason. 2017. “Code of Silence The Docket: National Pulse.” *ABA Journal* 103 (12): 18–19.
Thatcher, Jim, David O’Sullivan, and Dillon Mahmoudi. 2016. “Data Colonialism through Accumulation by Dispossession: New Metaphors for Daily Data.” *Environment and Planning D: Society and Space* 34 (6): 990–1006.
The Washington Post. 2018. “Transcript of Mark Zuckerberg’s Senate Hearing.” *The Washington Post*, April 10, 2018. https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/?noredirect=on.
Turkel, Gerald. 1988. “The Public/Private Distinction: Approaches to the Critique of Legal Ideology.” *Law & Society Review* 22: 801–28.
UW Data Science for Social Good Program. 2017. “Improving Transit Services Using ORCA Data,” 2017. https://escience.washington.edu/dssg/project-summaries-2017/.
Varian, Hal R., Joseph Farrell, and Carl Shapiro. 2004. *The Economics of Information Technology: An Introduction*. Cambridge University Press.
Volokh, Eugene. 2000. “Freedom of Speech and Information Privacy: The Troubling Implications of a Right to Stop People from Speaking about You.” *Stanford Law Review*, 1049–1124.
Wamsley, Gary L., and Mayer N. Zald. 1973. “The Political Economy of Public Organizations.” *Public Administration Review*, 62–73.
Warzel, C. 2015. “Let’s All Join the AP Stylebook in Killing the Term Ride-Sharing.” *Buzzfeed News*, January 8, 2015. https://www.buzzfeednews.com/article/charliewarzel/lets-all-join-the-ap-stylebook-in-killing-the-term-ride-shar#.ksk27KmN0.
Washington State Auditor’s Office. 2016. “Performance Audit: The Effect of Public Records Requests on State and Local Governments.” https://sao.wa.gov/the-cost-of-transparency/.
Wedel, Janine R., Cris Shore, Gregory Feldman, and Stacy Lathrop. 2005. “Toward an Anthropology of Public Policy.” *The ANNALS of the American Academy of Political and Social Science* 600 (1): 30–51. https://doi.org/10.1177/0002716205276734.
Weintraub, Jeff. 1997. “The Theory and Politics of the Public/Private Distinction.” *Public and Private in Thought and Practice: Perspectives on a Grand Dichotomy* 1: 7.
Weintraub, Jeff, and Krishan Kumar. 1997. *Public and Private in Thought and Practice: Perspectives on a Grand Dichotomy*. University of Chicago Press.
Weiss, Robert S. 1995. *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. Simon and Schuster.
Wexler, Rebecca. 2015. “Convicted by Code.” *Slate Magazine*, October 6, 2015. http://www.slate.com/blogs/future_tense/2015/10/06/defendants_should_be_able_to_inspect_software_code_used_in_forensics.html.
Whittington, Jan. 2008. “The Transaction Cost Economics of Highway Project Delivery: Design-Build Contracting in Three States.” University of California, Berkeley.
———. 2012. “When to Partner for Public Infrastructure? Transaction Cost Evaluation of Design-Build Delivery.” *Journal of the American Planning Association* 78 (3): 269–285.
Whittington, Jan, Ryan Calo, Mike Simon, Jesse Woo, Meg Young, and Peter Schmiedeskamp. 2016. “Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government.” *Berkeley Technology Law Journal* 30 (3).

http://papers.ssrn.com/abstract=2636074.
Williamson, Oliver E. 2002. “The Theory of the Firm as Governance Structure: From Choice to Contract.” *Journal of Economic Perspectives* 16 (3): 171–195.
Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Oxford: Blackwell.
Yergin, Daniel, and Joseph Stanislaw. 2002. *The Commanding Heights: The Battle for the World Economy*. Simon and Schuster.
Young, Meg, Michael Katell, and P. M. Krafft. 2019. “Municipal Surveillance Regulation and Algorithmic Accountability.” *Big Data & Society* 6 (2): 2053951719868492.
Young, Meg, Luke Rodriguez, Emily Keller, Feiyang Sun, Boyang Sa, Jan Whittington, and Bill Howe. 2019. “Beyond Open vs. Closed: Balancing Individual Privacy and Public Accountability in Data Sharing.” In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 191–200.
Yu, H., and D. Robinson. 2012. “The New Ambiguity of ‘Open Government’. Princeton CITP.” Yale ISP Working Paper.
Zenner, A. 2015. “The AP Bans the Term ‘Ride-Sharing’ for Uber and Lyft.” *Greater Greater Washington, Jan* 14.
Zhang, Ying, Tom Thomas, M. J. G. Brussel, and MFAM Van Maarseveen. 2016. “Expanding Bicycle-Sharing Systems: Lessons Learnt from an Analysis of Usage.” *PLoS One* 11 (12): e0168604.

APPENDIX A: EXAMPLE INTERVIEW PROTOCOL

King County Metro Interview Protocol

A. INTRO

1. What is your full job title?

2. What are your job duties?

3. Will you please walk me through a typical day in your role?

B. DETAILS ON DESIGN AND FUNCTIONING

1. I am interested in learning more about the Mobility on Demand program, will you please describe to me what it is, in detail?

2. What data has been requested so far from the firm/Lyft?

3. What is the department's motivation to collect this data?

4. How would the department use this data?

5. What is your role with respect to this program?

6. What was the initial catalyst for this program?

7. What mechanisms does KCM have to compel the firm to share data? Have they used them?

8. What has been the firm's/Lyft's response to date?

9. What factors affect whether you can get access to the data you are looking for?

10. Who owns the data that KCM would get from a firm/Lyft?

C. DECISION PROCESS OF SELECTING A PARTICULAR VENDOR SYSTEM

1. I am interested in understanding the context for the decisions KCM made in selecting a vendor for this system.

2. Would I be able to see documents submitted for the MOD Sandbox application?

3. What is the long-term vision for this product/service?

D. DEGREE OF AGENCY INPUT ON SYSTEM DESIGN AND FUNCTIONALITY

1. What input is the firm giving the agency into what data they collect?

E. INTRO TO:

1. Public records person
    - How public records requests of vendor data are handled
2. Contracts person
    - Terms of the vendor agreement
3. Anyone involved in application
4. Anyone involved in data bargaining
5. Data analysts using TNC data
6. Anyone at Lyft/firm

Misc.

Software functionality
Pricing model
How privacy is protected in the design and execution of such data systems
How companies decide what data is proprietary
Is the data valuable for any purpose other than sharing it with public agencies?
What processing, if any, does your data go through prior to being shared with Sound Transit/ORCA?

APPENDIX B: LANGUAGE OF PUBLIC RECORDS REQUESTS

Subject: Washington Public Records Act Request: Internal ORCA/ST documents related to ORCA data ecosystem

Email

To Whom It May Concern:

Pursuant to the Washington Public Records Act, I hereby request the following records:

1. Copy of an ORCA organization "org" chart that includes the 7 partner agency Site Managers.

2. A copy of every formal "change request" for the vendor agreement between ORCA and Vix between 2016 - 2019 (formerly AIG).

3. A copy of every signed version of the vendor agreement between ORCA and Vix (formerly AIG); in the event there are many versions of the vendor agreement, I would be grateful for only versions that were signed between 2016-2019.

4. Original RFP for the original ORCA card system.

5. A copy of any 5 law enforcement requests for ORCA data related to a court order or subpoena.

6. A copy of any 5 institutional ORCA customers/users requests for ORCA data or reporting (e.g. as Amazon, Microsoft, University of Washington, QFC).

7. Copies of all public records requests made in 2018 related to the ORCA card program.

8. All emails related to the 2018 public records request for the software application for the portable handheld fare enforcement devices.

I also request that, if appropriate, fees be waived as I believe this request is for my PhD research at the University of Washington. These documents are to support to my PhD dissertation research in which I ask, "How do public transportation agencies work with private firms to enable policy programs that rely on continuous or geospatial data? What contextual and historical factors led to current system design and institutional configurations? How do stakeholders define the roles of their respective organizations?" More information on my dissertation is here: https://sites.uw.edu/dataintransit/

In the event that fees cannot be waived, I would be grateful if you would inform me of the total charges in advance of fulfilling my request. I would prefer the request filled electronically, by e-mail attachment if available or CD-ROM if not.

Thank you in advance for your help. I look forward to receiving your response to this request and appreciate your time on this request.

Sincerely,
Meg Young

Subject: Washington Public Records Act Request: Agency access to data and finding what is considered proprietary

Email

To Whom It May Concern:

Pursuant to the Washington Public Records Act, I hereby request the following records:

1. User manuals or documentation of SAP Business Objects and/or SAS used internally for managing ORCA data
2. User manuals or documentation related to ORCA card hardware owned by ORCA (e.g. validators)
3. Agency policy, guidance, or vendor agreements that address data sharing and interoperability between Vix and the makers of the ORCA card vending machines (Schiedt & Bachman)
4. Document listing names of "universes" and "fields" available in the SAP Business Objects application
5. Emails between employees of Sound Transit/ORCA and Vix regarding agency access to data for the Data Access and Reporting warehousing effort.
6. Requests for proposals, proposals submitted by vendors, and contracts, for the Next Gen ORCA system.
7. Any Sound Transit or ORCA retained documents describing the functioning of the "back office" reporting system, which has been described to me as owned by ORCA.

I also request that, if appropriate, fees be waived as I believe this request is for my PhD research at the University of Washington. These documents are to support to my PhD dissertation research in which I ask, "How do public transportation agencies work with private firms to enable policy programs that rely on continuous or geospatial data? What contextual and historical factors led to current system design and institutional configurations? How do stakeholders define the roles of their respective organizations?" More information on my dissertation is here: https://sites.uw.edu/dataintransit/

In the event that fees cannot be waived, I would be grateful if you would inform me of the total charges in advance of fulfilling my request. I would prefer the request filled electronically, by e-mail attachment if available or CD-ROM if not.

Thank you in advance for your help! I look forward to receiving your response to this request and appreciate your time and help very much.

Sincerely,
Meg Young
Data Lab, UW Information School

Subject: Washington Public Records Act Request: Agency access to partner data & what is considered proprietary

To Whom It May Concern:

Pursuant to the Washington Public Records Act, I hereby request the following records:

1. Requests for proposals, proposals submitted by vendors, and draft contracts, for the MOD Sandbox program between Lyft and KCM
2. Requests for proposals, proposals submitted by vendors, and contracts, for the MOD Sandbox program between Via and KCM
3. Draft data sharing agreements between Lyft and KCM
4. Draft and final data sharing agreements between Via and KCM
5. The final 1-2 emails that Lyft employees sent to KCM regarding the MOD Sandbox program

I also request that, if appropriate, fees be waived as this request is for my PhD research at the University of Washington. These documents are to support to my PhD dissertation research in which I ask, "How do public transportation agencies work with private firms to enable policy programs that rely on continuous or geospatial data? What contextual and historical factors led to current system design and institutional configurations? How do stakeholders define the roles of their respective organizations?" More information on my dissertation is here: https://sites.uw.edu/dataintransit/

In the event that fees cannot be waived, I would be grateful if you would inform me of the total charges in advance of fulfilling my request. I would prefer the request filled electronically, by e-mail attachment if available, or CD-ROM if not.

Thank you in advance for your help! I look forward to receiving your response to this request and appreciate your time and help very much.

Sincerely,

Meg Young

Subject: California Public Records Act Request: MOD Sandbox Program (Metropolitan Transportation Authority)

To Whom It May Concern:

Pursuant to the California Public Records Act, I hereby request the following records:

Any 2017-2018 emails in which LA Metro, King County Metro, or Lyft employees state the outcome of the MOD Sandbox negotiations between KCM, LA Metro, and Lyft, especially any records officially ending the partnership and which state the reason(s) for dissolving the FTA MOD Sandbox pilot with Lyft.

I am writing a case study for school.

Thank you!

The requested documents will be made available to the general public, and this request is not being made for commercial purposes.

In the event that there are fees, I would be grateful if you would inform me of the total charges in advance of fulfilling my request. I would prefer the request filled electronically, by e-mail attachment if available or CD-ROM if not.

Thank you in advance for your anticipated cooperation in this matter. I look forward to receiving your response to this request within 10 calendar days, as the statute requires.

Sincerely,

Meg Young

APPENDIX C: EXCERPTS FROM ORIGINAL ORCA CONTRACT

CONTRACT for the

DEVELOPMENT, IMPLEMENTATION, OPERATION AND MAINTENANCE

of the
REGIONAL FARE COORDINATION SYSTEM

Division I: Contract Terms and Conditions Division II: Services Specifications
Division III: Equipment Specifications

CONTRACT NO. 229944

35.7 Ownership of Use Data and Use Data Reports

35.7.1 Agency Ownership

(a) All Use Data, Use Data Reports and all rights thereto, shall be the sole and exclusive property of the Agencies. Contractor hereby irrevocably assigns exclusively to the Agencies and their successors and assigns any and all right, title and interest in the Use Data, including all copyrights, trade secret rights, and other proprietary rights, title, and interest thereto. For clarification, any Use Data Reports that are created by the Agencies shall be the property of the Agencies. To the extent it may be deemed by operation of law at any time that the Agencies are not the sole owners of all possible rights in and to any of the Use Data or Use Data Reports, or that Contractor, its members, or its Subcontractors retain any rights to the same other than those provided in this Contract, Contractor hereby irrevocably grants to the Agencies and their successors and assigns the exclusive unrestricted right in perpetuity to use the same and exercise all such rights on a royalty-free, worldwide, fully transferable basis. To the extent it may be deemed that any assignment or grant of rights under this paragraph cannot be made until after the relevant Use Data is in existence, Contractor's acceptance of any payment under this Contract shall constitute such an assignment or grant with respect to all such complete or incomplete Use Data that exists as of the date such payment is accepted. Contractor agrees to execute at any time such documents as may be requested by the Agencies to evidence or perfect such assignment or the Agencies' proprietary and intellectual property rights as stated in this paragraph, but the Agencies' failure to request the execution of such documentation shall not affect the existence of the Agencies' rights as stated in this subsection.

(b) The Agencies grant the Contractor a non-exclusive, royalty-free license to use the Use Data for purposes relating to the performance by the Contractor of any of its obligations under the Contract.

35.7.2 Privacy and Confidentiality of Use Data and Use Data Reports

The Contractor shall at all times maintain the privacy and confidentiality of the Use Data and any Use Data Reports created by the Contractor or its Subcontractors. All Use Data and Use Data Reports shall be at all times segregated and kept confidential by the Contractor, and may not be used or disclosed, in whole or in part, in any manner except as expressly authorized by this Contract or with the written consent of the Contract Administrator. All Use Data and Use Data Reports shall be returned to the Agencies upon request or, if earlier, upon expiration or termination of this Contract. To the extent that such materials consist of copies of information also in the Agencies' possession, Contractor may alternately certify in writing to the Agencies that the materials have been destroyed.

APPENDIX D: AMENDMENT PROVIDING FOR ORCA DATA ACCESS

Amendment 400
Contract No. 229944

To the Contract for the Design, Implementation, Operation and Maintenance of the Regional Fare Coordination System

This Amendment 400 to the Contract for the Design, Implementation, Operation and Maintenance of the Regional Fare Coordination System is entered into this 9TH day of AUGUST, 2017, by and between Vix Technology (USA) Inc (formerly known as ERG Transit Systems (USA Inc), a California corporation and wholly owned subsidiary of Vix Mobility Pty Ltd, an Australian corporation, (hereinafter referred to as the "Contractor") and each of the following seven public transportation agencies (hereinafter referred to individually as an "Agency" or collectively as the "Agencies"):

1. Central Puget Sound Regional Transit Authority ("Sound Transit")
2. King County ("King County")
3. Kitsap County Public Transportation Benefit Area ("Kitsap Transit")
4. Pierce County Public Transportation Benefit Area ("Pierce Transit")
5. Snohomish County Public Transportation Benefit Area ("Community Transit")
6. City of Everett ("Everett")
7. State of Washington, acting through the Washington State Department of Transportation, Washington State Ferries Division ("WSF")

Recitals

A. Effective April 29, 2003, each of the Agencies and the Contractor entered into Contract #229944 ("Contract") to implement a Regional Fare Coordination System ("RFC System") to establish a common fare system utilizing smart card technology. The Contractor is responsible for the development, implementation, operation and maintenance of the RFC System as specified in the Contract.

B. The Agencies and the Contractor desire to amend Section VI of Exhibit 9, Price Schedule Special Programs, to compensate the Contractor for performing the development work necessary to facilitate the migration of ORCA usage data to an internal ORCA Agency data repository to help validate and drive ngORCA reporting requirements.This work is performed per *CR-13002 ngORCA Data Migration Support v6.0* as approved by the Agencies on July 19, 2017.

C. The Parties agree that the Work necessary to modify the ORCA system as directed will be performed and compensated as described below.

Agreement
Section 1.0 Description of Work

Summary

1.1 The Contractor will use Oracle GoldenGate to migrate and synchronize the data objects between databases.

1.2 Once fully implemented the Production Disaster Recovery (DR) site will be used as the source database.

Prerequisites

1.3 The Agencies will be responsible for setting up, configuring and maintaining Oracle GoldenGate on target database. This is required to retrieve data migrated from the source database to staging server.

1.4 The staging server will be a Linux node hosting Oracle RDBMS binaries (OEL/RHEL 5-UL6, 6-UL1, 7 or SLES 11).

1.5 The Agencies will stipulate what database product they intend to use on the target database.

1.6 The Contractor will verify that the selected database product will be compatible with the solution.

1.7 The current assumption is that Agencies will use an Azure hosted SQL Server.

1.8 The Agencies will be responsible for setting up, configuring and maintaining the Virtual Private Network (VPN) connection required to access the DMZ (Demilitarized Zone) from their network (hosting the target database).

1.9 The Contractor will provide support as needed to the Agencies to install and configure Oracle GoldenGate on the target database.

1.10 The Contractor will assist with other environmental configurations (i.e. firewall updates)

Testing

1.11 Prior to deployment, the Contractor shall configure an instance of Oracle GoldenGate to replicate data from Regional Test Bed to a staging RHEL (Linux server) in a DMZ and then to the (target) database to best mimic the production deployment.

1.12 This environment will be used to verify the configuration and implementation of the solution as described in production implementation (1.15 –1.33)

1.13 Once testing has been verified, Oracle GoldenGate will be completely de-installed from the test environment by the Contractor.

1.14 Post deployment, if the Agencies would like the Contractor to spin up a test environment a new change request must be submitted.

Production Implementation

1.15 The Contractor will migrate the following data sources from the Production Disaster Recovery site to a staging server.

Data Sources

- **CUT** (schema)
  - **CUT_PI_ENTRY** (table)
    - Contains Purse, pass and multiride use on entry transactions
  - **CUT_PI_EXIT** (table)
    - Contains Purse, pass and multiride use on entry transactions
  - **CUT_PI_FINANCIAL** (table)
    - Contains a wide range of product issuance, revalue, reversal, refund, recovery, write off and surcharge transactions
  - **CUT_PI_JOURNEY** (table)
    - Contains Purse, pass and multiride use journey transactions and product multiride rebate on exit transaction
- **BUSINESS_ACCOUNT** (schema)
  - **INSTITUTION** (table)
    - Contains institution details, such as name, address, contract dates
  - **INSTITUTION_AGREEMENT** (table)
    - Contains agreement details, such as (unique) IDs, addresses, lead agent contact details
  - **INSTITUTION_CARD_GROUP** (table)
    - Contains card group IDs and details (such as suspension dates) associated with institutions (institutional cards are associated to card groups belonging to institutions)
  - **INSTITUTION_CARD** (table)
    - Contains (institutional) card serial numbers associated with card groups
- **XWEB** (schema)
  - **CARD_DETAILS** (table)
    - Contains (all) card serial numbers and associated cardholder IDs (does not provide card verification numbers)
  - **CARDHOLDER_ACCOUNT_DETAILS** (table)
    - Contains all cardholder account details (i.e. Name, addresses, contact details and other PII data

1.16 The staging server will be located in the DMZ.

1.17 The Contractor will be responsible for the configuration and maintenance of the staging server.

1.18 Oracle GoldenGate will migrate data (in the form of trail files from the Production Disaster Recovery site Offline Server (OFS) to the staging server).

1.19 The Contractor will work collaboratively with the Agencies via iterative testing and development to support data source delivery incrementally.

1.20 The Agencies will be responsible for providing, configuring and maintaining the target database, and the Oracle GoldenGate instance that resides on it.

1.21 The Contractor will provide support to the Agencies in configuring the Oracle GoldenGate instance on the target database in order to access the Oracle GoldenGate files migrated to proxy server.

1.22 The Contractor will provide the Agencies the required GoldenGate licenses (Named User Plus) and will be responsible for the ownership of these licenses.

1.23 If the Agency target database has more than x1 processor (as determined by the Oracle Core Processing Factor) then additional licenses will be required.

1.24 Once transferred, the Agencies shall be responsible for all data contained on their (target) database. This includes any backups of data.

(a) The Contractor shall not be responsible for any data lost during transfer from the Production Disaster Recovery (DR) site.

(b) The Contractor shall not provide support once data has reached the target database.

1.25 The data sources in 1.15 do not provide Configuration Data (CD) context data.

(a) i.e. mapping of internal trip/route IDs to that provided in import sets.

1.26 None of the data sources referenced in 1.15 contain credit card details.

(a) However, the CARDHOLDER and BUSINESS_ACCOUNT data sources do contain business names, names of individuals, addresses, contact details and other PII data.

1.27 The Contractor can provide a size estimate of the data sources once the Agencies determine what data sources to migrate.

1.28 The Contractor will not be responsible for training Agency resources or Agency third party contractors in regards to data transferred to the target database outside of documentation to be provided as part of this body of work (i.e. ERD and data dictionary).

1.29 Only data currently available on the Offline Server (OFS) at the time Oracle GoldenGate is activated will be migrated to the target database

(a) Any purged data will not be migrated

1.30 Oracle GoldenGate will require installation on the source and target databases as well as the staging server.

1.31 Oracle EM Cloud Control 13c provides a plugin for Oracle GoldenGate for system monitoring.

1.32 The Agencies shall raise additional, separate change requests if they require any additional data sources

Optional CUT Schema Tables

1.33 The following nine optional CUT schema tables are available for an additional fixed package price of $34,290 this is in addition to the $250,862 solution price.

- **CUT** (schema)
  - **CUT_CI_BLOCKING** (table)
    - Contains card block and unblock transactions
  - **CUT_CARDHOLDER_FINANCIAL** (table)
    - Contains remote purchase and reversal transactions
  - **CUT_CARDHOLDER_FEE** (table)
    - Contains fee and remote fee (and reversal) transactions
  - **CUT_CI_FINANCIAL** (table)
    - Contains card fee and revenue allocation transactions
  - **CUT_FINANCIAL** (table)
    - Contains a wide range of financial transactions, such as institution order payments, other fund payments
  - **CUT_PI_BLOCK** (table)
    - Contains product block and unblock transactions
  - **CUT_PI_MAINTENANCE** (table)
    - Contains auxiliary product transactions, such as product delete, autoload update, replacement transactions
  - **CUT_RIDERSHIP_STATS** (table)
    - Contains ridership events transaction
  - **STREAMING_SESSION_HISTORY** (table)
    - Provides association of streaming session IDs to number of transactions streamed in that session, along with the session dates (start and end)

Documentation Updates

The Contractor will finalize and deliver documentation by February 28, 2018. Draft documentation will be made available prior to this date

(a) SEA-00TBD Data Migration Using Oracle GoldenGate (or similar)

Section 2.0 Schedule

2.1 The work described in Section 1.0 will be completed by December 31, 2017.

APPENDIX E: EXEMPTION IN PUBLIC RECORDS ACT FOR PROPRIETARY INFROMATION (RCW 42.56.270)

Financial, commercial, and proprietary information.

*** CHANGE IN 2020 *** (SEE 5549-S2.SL) ***

The following financial, commercial, and proprietary information is exempt from disclosure under this chapter:

(1) Valuable formulae, designs, drawings, computer source code or object code, and research data obtained by any agency within five years of the request for disclosure when disclosure would produce private gain and public loss;

(2) Financial information supplied by or on behalf of a person, firm, or corporation for the purpose of qualifying to submit a bid or proposal for (a) a ferry system construction or repair contract as required by RCW 47.60.680 through 47.60.750; (b) highway construction or improvement as required by RCW 47.28.070; or (c) alternative public works contracting procedures as required by RCW 39.10.200 through 39.10.905;

(3) Financial and commercial information and records supplied by private persons pertaining to export services provided under chapters 43.163 and 53.31 RCW, and by persons pertaining to export projects under RCW 43.23.035;

(4) Financial and commercial information and records supplied by businesses or individuals during application for loans or program services provided by chapters 43.325, 43.163, 43.160, 43.330, and 43.168 RCW, or during application for economic development loans or program services provided by any local agency;

(5) Financial information, business plans, examination reports, and any information produced or obtained in evaluating or examining a business and industrial development corporation organized or seeking certification under chapter 31.24 RCW;

(6) Financial and commercial information supplied to the state investment board by any person when the information relates to the investment of public trust or retirement funds and when disclosure would result in loss to such funds or in private loss to the providers of this information;

(7) Financial and valuable trade information under RCW 51.36.120;

(8) Financial, commercial, operations, and technical and research information and data submitted to or obtained by the clean Washington center in applications for, or delivery of, program services under *chapter 70.95H RCW;

(9) Financial and commercial information requested by the public stadium authority from any person or organization that leases or uses the stadium and exhibition center as defined in RCW 36.102.010;

(10)(a) Financial information, including but not limited to account numbers and values, and other identification numbers supplied by or on behalf of a person, firm, corporation, limited liability company, partnership, or other entity related to an application for a horse racing license submitted pursuant to RCW 67.16.260(1)(b), marijuana producer, processor, or retailer license, liquor license, gambling license, or lottery retail license;

(b) Internal control documents, independent auditors' reports and financial statements, and supporting documents: (i) Of house-banked social card game licensees required by the gambling commission pursuant to rules adopted under chapter 9.46 RCW; or (ii) submitted by tribes with an approved tribal/state compact for class III gaming;

(c) Valuable formulae or financial or proprietary commercial information records received during a consultative visit or while providing consultative services to a licensed marijuana business in accordance with RCW 69.50.561;

(11) Proprietary data, trade secrets, or other information that relates to: (a) A vendor's unique methods of conducting business; (b) data unique to the product or services of the vendor; or (c) determining prices or rates to be charged for services, submitted by any vendor to the department of social and health services or the health care authority for purposes of the development, acquisition, or implementation of state purchased health care as defined in RCW 41.05.011;

(12)(a) When supplied to and in the records of the department of commerce:

(i) Financial and proprietary information collected from any person and provided to the department of commerce pursuant to RCW 43.330.050(8); and

(ii) Financial or proprietary information collected from any person and provided to the department of commerce or the office of the governor in connection with the siting, recruitment, expansion, retention, or relocation of that person's business and until a siting decision is made, identifying information of any person supplying information under this subsection and the locations being considered for siting, relocation, or expansion of a business;

(b) When developed by the department of commerce based on information as described in (a)(i) of this subsection, any work product is not exempt from disclosure;

(c) For the purposes of this subsection, "siting decision" means the decision to acquire or not to acquire a site;

(d) If there is no written contact for a period of sixty days to the department of commerce from a person connected with siting, recruitment, expansion, retention, or relocation of that person's business, information described in (a)(ii) of this subsection will be available to the public under this chapter;

(13) Financial and proprietary information submitted to or obtained by the department of ecology or the authority created under chapter 70.95N RCW to implement chapter 70.95N RCW;

(14) Financial, commercial, operations, and technical and research information and data submitted to or obtained by the life sciences discovery fund authority in applications for, or delivery of, grants under **chapter 43.350 RCW, to the extent that such information, if revealed, would reasonably be expected to result in private loss to the providers of this information;

(15) Financial and commercial information provided as evidence to the department of licensing as required by RCW 19.112.110 or 19.112.120, except information disclosed in aggregate form that does not permit the identification of information related to individual fuel licensees;

(16) Any production records, mineral assessments, and trade secrets submitted by a permit holder, mine operator, or landowner to the department of natural resources under RCW 78.44.085;

(17)(a) Farm plans developed by conservation districts, unless permission to release the farm plan is granted by the landowner or operator who requested the plan, or the farm plan is used for the application or issuance of a permit;

(b) Farm plans developed under chapter 90.48 RCW and not under the federal clean water act, 33 U.S.C. Sec. 1251 et seq., are subject to RCW 42.56.610 and 90.64.190;

(18) Financial, commercial, operations, and technical and research information and data submitted to or obtained by a health sciences and services authority in applications for, or delivery of, grants under RCW 35.104.010 through 35.104.060, to the extent that such information, if revealed, would reasonably be expected to result in private loss to providers of this information;

(19) Information gathered under chapter 19.85 RCW or RCW 34.05.328 that can be identified to a particular business;

(20) Financial and commercial information submitted to or obtained by the University of Washington, other than information the university is required to disclose under RCW 28B.20.150, when the information relates to investments in private funds, to the extent that such information, if revealed, would reasonably be expected to result in loss to the University of Washington consolidated endowment fund or to result in private loss to the providers of this information;

(21) Market share data submitted by a manufacturer under RCW 70.95N.190(4);

(22) Financial information supplied to the department of financial institutions, when filed by or on behalf of an issuer of securities for the purpose of obtaining the exemption from state securities registration for small securities offerings provided under RCW 21.20.880 or when filed by or on behalf of an investor for the purpose of purchasing such securities;

(23) Unaggregated or individual notices of a transfer of crude oil that is financial, proprietary, or commercial information, submitted to the department of ecology pursuant to RCW 90.56.565(1)(a), and that is in the possession of the department of ecology or any entity with which the department of ecology has shared the notice pursuant to RCW 90.56.565;

(24) Financial institution and retirement account information, and building security plan information, supplied to the liquor and cannabis board pursuant to RCW 69.50.325, 69.50.331, 69.50.342, and 69.50.345, when filed by or on behalf of a licensee or prospective licensee for the purpose of obtaining, maintaining, or renewing a license to produce, process, transport, or sell marijuana as allowed under chapter 69.50 RCW;

(25) Marijuana transport information, vehicle and driver identification data, and account numbers or unique access identifiers issued to private entities for traceability system access, submitted by an individual or business to the liquor and cannabis board under the requirements of RCW 69.50.325, 69.50.331, 69.50.342, and 69.50.345 for the purpose of marijuana product traceability. Disclosure to local, state, and federal officials is not considered public disclosure for purposes of this section;

(26) Financial and commercial information submitted to or obtained by the retirement board of any city that is responsible for the management of an employees' retirement system pursuant to the authority of chapter 35.39 RCW, when the information relates to investments in private funds, to the extent that such information, if revealed, would reasonably be expected to result in loss to the retirement fund or to result in private loss to the providers of this information except that (a) the names and commitment amounts of the private funds in which retirement funds are invested and

(b) the aggregate quarterly performance results for a retirement fund's portfolio of investments in such funds are subject to disclosure;

(27) Proprietary financial, commercial, operations, and technical and research information and data submitted to or obtained by the liquor and cannabis board in applications for marijuana research licenses under RCW 69.50.372, or in reports submitted by marijuana research licensees in accordance with rules adopted by the liquor and cannabis board under RCW 69.50.372;

(28) Trade secrets, technology, proprietary information, and financial considerations contained in any agreements or contracts, entered into by a licensed marijuana business under RCW 69.50.395, which may be submitted to or obtained by the state liquor and cannabis board;

(29) Financial, commercial, operations, and technical and research information and data submitted to or obtained by the Andy Hill cancer research endowment program in applications for, or delivery of, grants under chapter 43.348 RCW, to the extent that such information, if revealed, would reasonably be expected to result in private loss to providers of this information;

(30) Proprietary information filed with the department of health under chapter 69.48 RCW; and

(31) Records filed with the department of ecology under chapter 70.375 RCW that a court has determined are confidential valuable commercial information under RCW 70.375.130.

VITA

Meg Young applies ethnographic and design methods to understand government use of information technologies on- the-ground, with a focus on how to make proprietary systems more accountable to the public. Her work to date reports on fieldwork with public agencies, advocates, and activists on data governance, surveillance, privacy, open records, data ownership, public-private partnerships, public engagement, and data trusts. This work has been published in ACM FACCT*, Big Data & Society, CHI, Ethics & Information Technology, and the Berkeley Technology and Law Journal. She is the co-founder of the Critical Platform Studies Group, a non-profit that partners with civil rights groups to pursue algorithmic accountability through participatory design.

After being awarded her PhD at the University of Washington Information School, she is joining Cornell Tech's Digital Life Initiative in New York City as a Postdoctoral Fellow. She was previously a Research Associate at Jesus College at the University of Cambridge and has an MSc in Information Science and a BA in Cultural Anthropology from the University of Michigan in Ann Arbor.