# Privacy in the Smart City:
# Implications of Sensor Network Design, Law, and Policy for Locational Privacy

Melissa Gaughan

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Urban Planning

University of Washington

2016

Committee:

Jan Whittington

Mark Purcell

Program Authorized to Offer Degree:

Urban Design and Planning

# Abstract

Automated data collection technology has transformed the city. From sensors that measure traffic flow to scanners that help police officers locate stolen vehicles, the use of smart city sensor networks has become an integral part of the modern city. Because the sensor networks collect highly detailed information about every aspect of city life, data from the sensor networks is often linked to individuals and can potentially be used to develop detailed profiles on residents. This thesis defines and identifies threats to locational privacy and identifies three major approaches to mitigating privacy risk: the use of tiered data structures modeled on IRB reviews; the use of privacy by design principles to minimize privacy-adverse data collection; and amendments to policy and law that would help to create a regulatory environment favorable to individual and societal privacy protection. This thesis involves case study research and interviews with program managers at the City of Seattle Policy Department, the City of Seattle Department of Transportation, and Sound Transit. It considers the locational privacy implications of automated license plate readers, WiFi beacon sniffer networks, and RFID transit passes. It concludes that locational privacy protection will require a careful blend of policy, law, and system design that is tailored both to the technology and to the larger political and legal environment in which the sensor network operates.

# Contents



# Tables and Figures

# Glossary

**ALPR:** Automated License Plate Readers

**FCC:** Federal Communications Commission

**FIPP:** Fair Information Privacy Principles

**FOIA:** Freedom of Information Act

**FTC:** Federal Trade Commission

**ICT:** Information and Communication Technologies

**MAC:** Media Access Control Device

**PBD:** privacy by design

**PII:** Personally Identifiable Information

**PRA:** Public Records Act

**RFID:** Radio Frequency Identification

**Reidentification:** The practice of re-associating personal information with anonymized records.

**Sensor Networks:** Data collection equipment used to collect location-based information continuously.

# 1. Introduction

In recent years, the debate about privacy in the information age has reached fever pitch. Within the first three months of 2016 alone, the FBI has demanded that Apple build software to bypass encryption on iPhones[1] and the federal government has pledged over 130 million dollars to funding smart city projects.[2] The perils and promises of Information and Communication Technologies (ICT) are almost without limit; like any other tool, the data gathered through new smart city programs can be used to make cities more just, equitable, and sustainable, or it can be used to strengthen the interests of the powerful and to marginalize vulnerable communities. As cities begin to develop new smart city data collection programs, it is crucial that decision makers weigh the risks and benefits of data collection and create policies to help ensure that data is collected in ways that respect personal privacy.

Both the ways in which data is collected and the *types* of data that are collected are rapidly shifting. With the rise of smart city sensor programs, which are made up of networks of sensors placed throughout the city that transmit data to the city's databases and operation centers in real time, cities are able to collect real-time information on electricity usage, traffic flows, pollution, and water consumption, among other things. In a complete smart city, almost all aspects of urban life can be turned into constant data flows captured by ubiquitous cameras and sensors placed throughout the urban landscape. The data flows can be cheaply aggregated, stored, and analyzed to draw conclusions both about city processes and about individual

[1] Apple, Inc. "A Message to Our Customers." Apple. Last modified February 16, 2016. Accessed March 22, 2016. http://www.apple.com/customer-letter/, 1.
[2] Office of the Press Secretary. "Obama Administration Announces New Smart City Initiatives." The White House. Last modified September 14, 2015. Accessed March 22, 2016. https://www.whitehouse.gov/the-press-office/2015/09/14/fact-sheet-administration-announces-new-smart-cities-initiative-help, 2.

residents. If done well, ubiquitous computing will enable cities to create real-time data-driven policies to mitigate problems and reduce resource waste.

While the benefits of smart city programs are impressive, these programs also have the potential to violate residents' privacy. After all, many of the data flows like traffic congestion, electricity consumption, and water usage are generated by individual residents. Research has shown that even anonymized public datasets can be used to identify a single individual and that the sophistication of reidentification techniques is keeping pace with new anonymization methods.[3] Data collection and dissemination is also problematic because it is almost impossible for individuals to track what personal information is disclosed and who has access to it. Attempting to manage one's own privacy can be like trying to track a particular dollar bill as it changes hands through hundreds of economic transactions: virtually impossible. Asking people to manage their own privacy is a proposition likely to fail.

Smart city sensor programs require a rethinking of privacy protection methods because they are markedly different from typical municipal data collection programs. Typical municipal data collection programs differ in three major ways from sensor networks: 1) data collection occurs on a periodic basis; 2) city agencies notify residents of data collection and obtain consent before data collection occurs; and 3) all data is either anonymized or aggregated before it is released to the public.[5] Typical data collection programs use notice and consent policies and back-end dataset anonymization to protect personal privacy. However, this methodology breaks down in the face of sensor technology, sophisticated reidentification techniques, and predictive

[3] McDonagh, Maeve. 2009. "The protection of personal information in public registers: the case of urban planning information in Ireland." Information & Communications Technology Law 18, no. 1: 19-38. Communication & Mass Media Complete, EBSCOhost (accessed December 2, 2015), 3.
[5] Scipioni, Marcello, Paolo, Dey, Anind, K., Chu, Hao-Hua, and Hayes, Gillian. "A Privacy-by-design Approach to Location Sharing." *Ubiquitous Computing Proceedings of the 2012 ACM Conference*, 2012, 580-83, 581.

analytics. Notice and consent are typically not automated features, and sensor systems may be installed without a commitment to deliver notice and obtain consent before collecting data, so oftentimes residents are unaware that they are being tracked as they move through the urban landscape.[6]

It is especially important that urban planners begin to think about both the risks and benefits of smart city sensor programs because planners are instrumental in shaping the future of their cities. In the past several years, planners have started to realize that technology has the potential to solve intransigent problems through the collection and use of data. By creating strategic plans, comprehensive plans, and operating budgets that specifically include smart city technologies, planners act as boosters for the implementation of new technologies in cities. While it is important that planners look to the future and recognize the role of technology in the city, it is also important to recognize that sensor network systems are not without risk. Because sensor networks have the ability to collect continuous flows of highly detailed information, and because most data collected by public agencies can be requested through local or federal public records requests, ensuring that data collected through smart city sensor networks cannot be used to track or target individuals or groups in a way that violates their privacy is a challenge that planners must fully consider before embracing smart city sensor networks as part of the future of the city.

Planners are well-suited to answering some of the fundamental questions of privacy in the smart city because they are versed in the various forms of notice and consent. Notice and consent are fundamental planning principles for everything from permitting to comprehensive

[6] Whittington, Jan, Ryan Calo, Mike Simon, Jesse Woo, Meg Young, and Peter Schmiedeskamp. "Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government." *Berkeley Technology Law Review* 30, no. 3 (May 4, 2015): 1899-1966, 1941.

planning outreach processes. Community support and buy-in are critical pieces of any planning process; there is no reason why the use of sensor networks to collect data should be any different. Planners can be champions for residents' rights to privacy and can give voice to concerns that might otherwise go unnoticed.

Given the potential rewards and risks of smart city sensor networks, it is time to reevaluate ways in which privacy can be protected in municipal data collection programs. This work focuses on protecting locational privacy in the smart city, which can be defined as the right to be free from unwanted surveillance in public spaces. It will seek to answer the following questions through a multi-part case study focusing on municipal agencies in the City of Seattle and at Sound Transit, the regional transportation agency:

1. What are the perils and promises of location-based smart city data collection programs?
2. What are cities with smart city data collection programs currently doing to protect locational privacy?
3. How can current/existing practices for protecting locational privacy be improved?

New insight into the protection of locational privacy is sought through a thorough review of privacy scholarship and through in-person interviews with city officials who are currently operating or developing smart city programs. The rest of this thesis is broken into chapters. Chapter Two is a review of existing literature, Chapter Three is a description of interview methods, Chapter Four will report on the findings from the interviews, Chapter Five will discuss strengths and weaknesses of current locational privacy protection methods, and Chapter Six will suggest potential improvements and propose directions for further research.

## 2. Literature Review

Privacy is a large and ever-growing field of scholarly pursuit. Privacy scholarship incorporates law, philosophy, economics, technology, ethics, psychology, sociology, computer science, data analytics, and security, just to name a few of the fields that have produced sizable bodies of literature on privacy.[7] Privacy has been dissected from many angles; scholars have developed descriptive and normative theories of privacy and have claimed both that privacy is a fundamental good and a good derived from other sources.[8,9] Some have gone so far as to argue that the value that we place on privacy is simply a product of our culture and not actually part of the human good at all.[10] Given the sprawling, nebulous, and tangled complexities of this issue, writing a coherent and compelling literature review has the potential to be a Sisyphean task.

However, building an understanding of informational privacy and the different varieties of privacy harms is crucial to this research. Scholars have noted that privacy suffers from a multitude of meanings; because the word "privacy" is used to justify so many different claims, it has become difficult to articulate a precise definition of privacy or to contextualize privacy in relation to transparency and freedom of speech. Luckily, Daniel J. Solove has created a taxonomy of privacy in an attempt to address this problem.[11] This

---

[7] Van den Hoven, Jeroen, Martijn Blaauw, Wolter Pieters, and Martijn Warnier. "Privacy and Information Technology." In Stanford Encyclopedia of Philosophy. Palo Alto, CA: Stanford University, 2014, 1.
[8] DeCew, Judith Wagner. In Pursuit of Privacy : Law, Ethics, and the Rise of Technology. Ithaca: Cornell University Press, 1997, 65.
[9] Moore, Adam D., and Moore, Adam D. Privacy Rights Moral and Legal Foundations. University Park, Pa.: Pennsylvania State University Press, 2010, 14.
[10] Finch 2014, 1598.
11 Solove, Daniel J. "A Taxonomy of Privacy." University of Pennsylvania Law Review 154, no. 3 (2006): 477-564.

literature review will use Solove's taxonomy of privacy to explain why privacy should be protected, and will then turn to the ethics and philosophies that can be used to justify a right to locational privacy; the ways in which privacy is currently protected through notice and consent regimes; and the possible solutions suggested by privacy scholars in multiple fields. This chapter seeks to answer the following questions:

1. What theories of privacy currently exist that can justify locational privacy?
2. What mechanisms are currently used to protect privacy? What are their strengths and weaknesses?
3. What theoretical constructs have been proposed as improvements to or replacements for the notice and consent privacy policy?

Although this literature review must by necessity remain narrow in scope, it will weave together disparate strands from many of the fields mentioned above in pursuit of a clear understanding of what work remains to be done both to articulate a compelling justification for locational privacy and to craft policies that are robust enough to protect the locational privacy of residents in a smart city.

## Taxonomy of Privacy

One of the major challenges in any work concerning privacy is articulating why it is valuable and what constitutes a privacy harm. For Daniel Solove, privacy is valuable because it is a relief from social pressures. The right to privacy gives people the freedom to pursue activities that might be countercultural or difficult to pursue in the glare of public scrutiny.[13] Privacy provides space to develop into unique beings; by allowing individuals to keep information private, societies allow space for diversity, curiosity, and development.

[13] Solove 2006, 8.

The right to privacy is not purely an individualistic right; because privacy promotes the social good and allows for diversity and development, privacy is a social good. Thus, a challenge to privacy is not simply an impingement of one person's right; it is a challenge to the social fabric.[14] Because privacy has important social and personal implications, it should be balanced with concerns like transparency and freedom of speech.

If privacy is to be weighed against concerns like transparency and freedom of speech, it is important to understand the different types of harms that can result from information collection, processing, and dissemination. Solove's taxonomy discusses all of the potential harms that can result from breaches of informational privacy. Figure 1 shows the relationship between data subjects and data holders, and names the categories of harms that can emerge from privacy breaches.

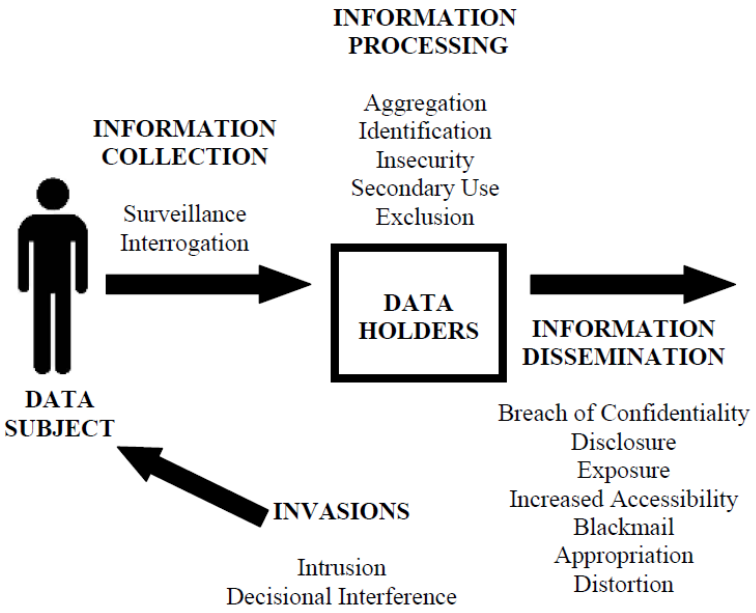
Privacy violations can result in several different types of harms. The two major types of harms are dignitary, where one person's standing, image, or self-worth is harmed through the non-consensual release of personal information. Architectural privacy harms, on the other hand, occur when there is a risk that a person or group of people might be harmed through the release of information sometime in the future. While this harm is more difficult to articulate, it bears similarities to environmental harms like pollution; while the threat of pollution may not be evident in any one solitary instance, the impacts of pollution are clear when studied at a global scale. Architectural harms have to do with situations where the data collection, use, and dissemination create long-term privacy risks for large

14 Solove 2006, 11.

populations in ways that cannot be fully comprehended until the harms have occurred. Most of the harms considered in this thesis are architectural in nature.

Solove's taxonomy of privacy is comprehensive, but for the sake of this literature review, I will address only a select number of Solove's harms. While all harms deserve to be considered, some are more relevant to locational privacy and municipal data collection programs. The following sections discuss potential harms that can result from information collection, information processing, information distribution, and privacy invasions as conceptualized in Solove's taxonomy of privacy.

*Figure 1 Taxonomy of Privacy*[15]




15 Solove 2006, 14.

Information collection involves activities that have to do with the gathering of information. The data collection programs that are considered in this thesis fall in to the category of surveillance, which is the watching, listening to, or recording of an individual's activities.[16] While surveillance is not necessarily problematic in all instances, the fear or awareness of constant surveillance can create a "Panopticon effect" and can chill socially transgressive actions. That is, surveillance can create conditions where individuals feel the need to self-regulate behavior, thereby limiting individual expression and freedoms.[17]

The second category, information processing, involves the way information is stored, manipulated and used. This, along with information dissemination, is perhaps the most important category of potential harms, because it is at this point that data is contextualized and combined to create information. Information processing involves both aggregation and identification, which is the combining of various pieces of data and the linking of information to individuals. These become potentially harmful when the data is used for purposes other than that which it was collected for. These secondary uses pose problems for consent-based programs and can be especially problematic if data is sold to data brokers who use information collected for legitimate city purposes to create detailed dossiers on consumers. When people are excluded from the conversation and not given the right to know about the data that others have about them, information processing creates an architectural harm for all data subjects.

The third group of privacy harms involves the dissemination of information. While breaches of information can be disastrous, this thesis focuses on harms resulting from

[16] Solove 2006, 14.
[17] Solove 2006, 44.

instances of disclosure and appropriation. Disclosure is the revelation of truthful information about a person that impacts the way others judge her character. Appropriation is the use of a person's personal information to serve the aims and interests of another.[18] While information dissemination can increase transparency and ensure that government data collection programs are not operating in secrecy, the widespread disclosure and appropriation of person information can also have deeply problematic implications. For example, a 2014 report by the FTC, data brokers—companies that collect consumers' personal information and resell or share that information with others—rely on government data obtained through public records requests to build profiles on nearly every U.S. consumer.[19] This complex, secretive, and largely unregulated industry uses public records to enable targeted marketing specifically tied to each individual's interests and spending habits. The FTC report established that there is a strong market for government data that uses datasets in ways far removed from their intended purposes. Thus, information dissemination can cause architectural harms for almost everyone in a society.

The fourth group of privacy harms are invasions into people's private affairs. Intrusion concerns invasive acts that disturb tranquility or solitude, while decisional interference is the government's influence on a data subject's decisions regarding her private affairs. While this thesis does not focus on invasions, it is important to note that intrusions and decisional interference are the end result of lax privacy practices.

[18] Solove 2006, 44.
[19] United States. Federal Trade Commission, Author. *Data Brokers : A Call for Transparency and Accountability*. Washington, D.C.]: Federal Trade Commission, 2014.

Solove's taxonomy of privacy is crucial for understanding the importance of informational privacy in the wider context of transparency and freedom of speech concerns because it demonstrates that informational privacy violations create both dignitary and architectural harms for individuals and for society as a whole. Privacy is a crucial element of a functioning society; it provides a space for self-reflection, independent action, and experimentation. As Hannah Arendt notes:

> A life spent entirely in public, in the presence of others, becomes, as we would say, shallow. While it retains its visibility, it loses the quality of rising into sight from some darker ground which must remain hidden if it is not to lose its depth in a very real, non-subjective sense.[20]

Privacy is valuable to both individuals and to societies generally. Solove has demonstrated that informational privacy results in harms, but work still remains to be done to examine how locational privacy can serve the interests of society and individuals. This next section examines the locational privacy's theoretical roots in informational privacy and privacy of the home.

## Origins of Locational Privacy

Like most philosophical concepts, locational privacy is a multi-faceted concept. Schools of thought on locational privacy abound, but for this literature review, I will limit the discussion of the origins of locational privacy to its origins in informational privacy and privacy of the home. Informational privacy is the idea that people should have the right to control access to personal information about themselves. Informational privacy is a key piece of online privacy, because the data that online transactions generate is a valuable

[20] Arendt, Hannah. The Human Condition. Charles R. Walgreen Foundation Lectures. Chicago: University of Chicago Press, 1958, 71.

asset in the information age. Since the early 1990's, the Federal Trade Commission (FTC) has been working to regulate privacy online and has focused mostly on protecting consumers' rights to control access and distribution of their personal information.[21] Solove's taxonomy of privacy primarily focuses on informational privacy; informational privacy is separate from physical privacy because informational privacy harms do not involve a physical harm. Instead, informational privacy violations create dignitary harms or architectural harms.

The concept of privacy of the home is a descendent of common law and is codified in the Fourth Amendment, which guarantees freedom from unreasonable searches and seizures and the Third Amendment, which guarantees that private homes will not be used to quarter soldiers. Courts have found that a person's reasonable expectation of privacy is dependent on their location. A man in his bedroom has a higher expectation of privacy than a man standing in the middle of a crowded music festival. Thus, courts deciding privacy cases have consistently found that people have a greater right to privacy in their homes than they otherwise would. For example, in a 2001 Supreme Court case, *Kyllo v. United States*, the court found that using thermal imaging technology to determine that the defendant was growing marijuana inside his home was a search that violated a reasonable expectation of privacy. If the defendant had been growing marijuana in a commercial warehouse, it is possible that he would have had no reasonable expectation of privacy. While spatial privacy is largely limited to variations on the "my home is my castle" argument right now, it is important to note that this argument is part of the intellectual

[21] The FTC will be covered at greater length in a later section.

foundation of locational privacy because locational privacy requires a right to privacy in public.

Informational privacy and privacy of the home are important to locational privacy because together they provide the conceptual foundation for a right to privacy in public, and by extension locational privacy. Because the ability to collect records of movement through space in permanent databases is relatively new, most theories of locational privacy rights draw strongly on informational privacy and privacy of the home for theoretical grounding.

## Locational Privacy

The study of locational privacy is still in its infancy; it is not protected by federal law,[22] nor is it considered in most privacy policies that have been adopted by cities. For the purposes of this paper, I will define locational privacy as the right to be free from unwanted systematic tracking in public spaces under normal circumstances. Most privacy policy, law, and scholarship are concerned with personally identifiable information (PII) in the form of names, birthdates, social security numbers, and account numbers. Locational privacy is a separate category of privacy research because it deals with people's movements through space and time. It is concerned with the ways in which travel patterns, address, or other geospatial data can be used to identify and track individuals. Even without access to a person's name or social security number, it is possible to uniquely

[22] "S. 2270: Location Privacy Protection Act of 2015." Govtrack.us. Last modified November 10, 2015. https://www.govtrack.us/congress/bills/114/s2270.

identify a person based on geospatial data points.[23] Traces of people in space and time have yet to be recognized as PII.

Because locational privacy is explicitly concerned with location, it is related to privacy of the home, as discussed above. Locational privacy is distinct from both informational privacy and privacy of the home because it combines elements of both theories of right in unique ways. The theory of locational privacy is both new and unique because it stems from recent developments in technologies that make it possible to convert ephemeral observations (i.e. noticing a few license plates while walking down the street) into a complete time stamped database of locational information.

Advances in technology have created new capabilities and generated a new need for a right to locational privacy that persists even in the public realm. Consider a city without ubiquitous sensors: collecting detailed information on a person's whereabouts would require hundreds of man-hours, which makes wide-scale location tracking prohibitively expensive.[24] In the age of ubiquitous computing and smart city sensor programs, granular data about residents' movements, actions, and transactions is becoming cheaper to collect.[25] Although smart city programs have positive implications for data-driven policy making, the new technology can also collect and retain detailed information about people through use of WiFi, Bluetooth, or RFID sensors, which can be used to develop detailed profiles on individuals or to profile marginalized communities.[26] As cities begin to enact smart city programs, it is important that policy makers think carefully about

[23] De Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. "Unique in the Crowd: The Privacy Bounds of Human Mobility." *Scientific Reports*3 (2013): Scientific Reports, 2013, Vol.3, 3.
[24] Finch and Tene 2014, 1588.
[25] Nissenbaum 2010, 114.
[26] Whittington et al 2016, 1931.

what data is being collected, how it is being gathered, and respect residents' right to locational privacy.

## Locational Privacy and the Right to Privacy in Public

The right to locational privacy can be derived from a theory of privacy in public. Helen Nissenbaum, a scholar of information privacy, is one of the strongest proponents of a right to privacy in public. Nissenbaum is an important and innovative scholar of informational privacy because she does not base her normative philosophy on a dichotomy between public and private spheres. While most other information privacy theorists argue that a right to privacy can only be justified in private spheres like the home or the doctor's office, Nissenbaum argues that privacy protection should be based on the context of the information exchange.[27]

For Nissenbaum, information norms are involved in any exchange of information, but the norms that are applicable are internal to a specific contract.[28] That is, norms are not universal. According to Nissenbaum:

> Contextual informational norms, like other social norms, generally, are not fixed and static, but may shift, fade, evolve and even reverse at varying rates, slowly or suddenly, sometimes due to deliberate cultural, legal, and societal alterations and other times in response to contingencies beyond human or societal control. Science and technology is a significant agent of change; in particular, computing and information technologies have been radically disruptive, enabling information practices that frequently diverge from entrenched informational norms.[29]

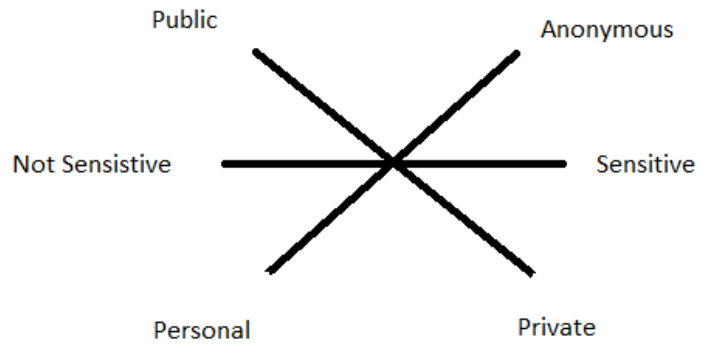
As the quote above demonstrates, Nissenbaum's theory of contextual integrity is rooted in already-developed norms of information sharing and control, but it is not meant to be a

---

[27] Nissenbaum, Helen Fay. *Privacy in Context : Technology, Policy, and the Integrity of Social Life*. Stanford, Calif.: Stanford Law Books, 2010, 132.
[28] Nissenbaum 2010, 132.
[29] Barocas, Solon, and Helen Nissenbaum. "Big Data's End Run around Anonymity and Consent." In *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, 44-75. Cambridge University Press, 2013, 47.

dogmatic framework of control. The degree to which information can rightly be shared is a result of the interplay between factors such as the relationship between the parties, the perceived sensitivity of the information, and whether or not information flows both ways (a casual conversation between friends) or only in one direction (a patient talking to her doctor about sensitive medical issues). All of these factors determine when information sharing (or collecting) are normal sharing of information or violations of informational privacy. The diagram below provides a preliminary sketch of the different paradigms of informational norms.

*Figure 2 Idea Space for Contextual Integrity*




The diagram above shows that contextual integrity is composed of grouped dichotomies. Information can be sensitive or not sensitive, public or private, personal or anonymous.[30] Information that is personal, private, or sensitive may require privacy protection, while information that is anonymous, public, or not sensitive is likely to not need privacy

[30] Nissenbaum, Helen. "Privacy as Contextual Integrity.(Symposium: Technology, Values, and the Justice System)." *Washington Law Review* 79, no. 1 (2004): 119-157, 138.

protection. This framework is domain neutral, meaning that the framework can be applied to many different spheres of information exchange. Informational norms prescribe information flows based on key actors, types of information, and constraints under which the information flow occurs.[31] That is, whether or not an exchange of information is appropriate depends on who is giving information, who is receiving that information, what they are talking about, and where they are talking. In other words, the ways in which these categories interact depends on the context in which the information is created. When a patient tells her doctor about her symptoms, it is appropriate that the flow of information only move in one direction (it is not appropriate for the doctor to share his medical history with the patient) and it is essential that the information is not distributed without the patient's consent. The fact that both the patient and the doctor abide by the norms of information flow and distribution make it possible for the patient to feel comfortable enough to share deeply personal information.

Because the norms that govern the appropriateness and flow of information stem from the domain of information exchange, it is possible to create a domain-specific framework for privacy protection. This is different from other theories of informational privacy because other theorists assert that there must be one consistent set of rules that govern all information exchanges if privacy is to be fully protected.[32] Nissenbaum's framework is more flexible because it allows for the norms of a particular context to dictate what is and is not a privacy violation. While this basis of analysis is much more

[31] Barocas, Solon, and Helen Nissenbaum. "Computing Ethics: Big Data's End Run Around Procedural Privacy Protections."*Association for Computing Machinery. Communications of the ACM 57*, no. 11 (2014), 46.

[32] Moore 2010, Regan 1995, and DeCew 1997.

nebulous than that used by scholars who base a right to privacy on the dichotomy between public and private spheres, Nissenbaum's contextual integrity is designed to protect informational privacy across many realms of information creation and exchange.

Contextual integrity provides a strong basis to a right to privacy in public because informational norms exist regardless of place. This makes contextual integrity a stronger basis for a right to privacy in public because it leaves more room for a sliding scale of privacy protection. While information gathered in the public sphere may not need the same degree of protection that medical records do, informational norms still apply in the public sphere. To put this more concretely, even though the public realm is a shared space where people go to see and be seen, there are behaviors that are not acceptable in that space. For example, American society has deemed it inappropriate for one person to stalk another because it violates personal privacy and is a nonconsensual activity. While the prohibition against stalking crosses between the public and private realms, it is equally inappropriate in both. There is no "public realm exception" that makes it socially acceptable for one person to follow another for long periods of time without consent.

Nissenbaum does acknowledge that the right to privacy in public seems counter-intuitive; if you choose to leave the sanctity of your home and enter the public realm, are you not giving up your right to privacy? To some extent, this claim is valid. However, being seen by people on the street is not the same thing as being recorded by ubiquitous computing sensor networks. In the era of big data, injudicious sensor programs can be used to collect so much location data that the location of each individual can be tracked as they

move through the city.[33] Furthermore, patterns of movement tend to be stable over time. That is, individuals make the same decisions regarding route, timing, and mode, which allows data scientists to predict where people will be throughout the day. These issues are particularly problematic from Nissenbaum's perspective because the data collection program is run by a government agency; spatial data collection programs are not a mutual sharing of information between two people. Instead, ubiquitous sensors collect data on everyone passing through a particular point in the city in order to develop a comprehensive and granular database of locational information. Just as interactions between individuals are governed by contextual norms, data collections programs should also abide by a set of contextual norms, agreed on by both the public and the data collectors.

The right to privacy in public is a strong theoretical justification for locational privacy because spatial data has clear informational norms associated with it. To apply Nissenbaum's framework to the locational data collection programs, the argument for privacy violation stems from the fact that the relationship between key actors (data subject and data collector) is both unequal and largely non-consensual; information flows only in one direction because most sensor programs do not have a mechanism for delivering notice to or obtaining consent from data subjects.[34] While the type of information collected is dictated by the data collection software, the constraints on information exchange are hotly contested because the information exchange occurs in the public realm. For all the reasons articulated above, I believe that entering the public realm does not mean an individual has

[33] Lane, Julia I. *Privacy, Big Data, and the Public Good: Frameworks for Engagement*. 2014, 10.
[34] Scipioni, Marcello, Paolo, Dey, Anind, K., Chu, Hao-Hua, and Hayes, Gillian. "A Privacy-by-design Approach to Location Sharing." *Ubiquitous Computing Proceedings of the 2012 ACM Conference*, 2012, 580-83.

ceded all right to privacy; even in the public realm, every private citizen should have the right to be free from unwanted pervasive surveillance.

When locational data is collected at the level of individual travelers, data collection crosses between public and private realms. For example, the majority of commute trips start or end at an individual's home. While data about home locations might not be directly collected by sensors, movement patterns are unique enough to uniquely identify individuals within large datasets.[35] Given that home ownership is a matter of public record, it would be easy to join a tax assessor's database to the trip information to build detailed profiles of where people live and work, which would violate a clear privacy norm.[36] When people are not made aware of these programs and given an opportunity to consent to data collection, there is a missed opportunity to build understanding about why data is being collected, how it is used, and the benefit the smart city programs can have for the residents themselves. Letting the data subjects have a voice in the design of the data collection program and building privacy protections into the data collection platform could prevent privacy breaches like the one described above. That is to say, public outreach and engagement can be used to shift informational norms in favor of smart city data collection programs, or curtail them in favor of the privacy preferences of residents.

Not all data collection programs in the public realm violate individuals' right to privacy; with proper policy and design, it is possible to collect data and increase operational efficiencies without collecting data that is tied to individuals. Nissenbaum

[35] Rossi, Luca, James Walker, and Mirco Musolesi. "Spatio-Temporal Techniques for User Identification by Means of GPS Mobility Data." 2015, 2.

[36] I am aware that contextual integrity leaves room for challenge on the grounds that the system is simply a glorification of preferences dressed up as cultural standards. This is a legitimate challenge to the argument; non-arbitrary methods of determining contextual norms will be discussed later in the literature review.

acknowledges that it is not always clear what information should be considered sensitive. For example, information like IP addresses, persistent unique identifiers, and locational data is not currently considered sensitive information in most privacy policies.[37] As the online environment changes and we become more aware of how information is gathered, aggregated, and analyzed, the norms of information exchange will evolve. By building privacy protection into data collection programs, cities can increase the level of trust that citizens feel for their governments, can avoid costly post-collection data redaction efforts, and can move towards an era of data-driven solutions.

As ubiquitous computing becomes more widespread and public realm surveillance technologies continue to improve, the need for a degree of privacy in the public realm becomes more and more evident. While most people are comfortable seeing others on the street, they may not be comfortable disclosing their names and final destinations to everyone they see. When surveyed by the Pew Research Forum[38], 88% of respondents said that it was important that they not be watched or listened to without their permission. This can clearly apply to smart city data collection programs that collect data in public space. As mentioned previously, recent research has shown that only four spatio-temporal data points are required to uniquely identify an individual within a city-level dataset.[39] There is a fundamental difference between a person being seen by other people on the street and individuals being tracked by in-situ sensors as they move through public space. The threat of architectural harms resulting from large scale data collection, processing, and

[37] Barocas and Nissenbaum 2014, 8.
[38] Pew Research Center. "American's Attitudes About Privacy, Security and Surveillance." Pew Research Center. Last modified March 20, 2015. Accessed March 22, 2016. http://www.pewinternet.org/files/2015/05/Privacy-and-Security-Attitudes-5.19.15_FINAL.pdf, 18.
[39] De Montjoye 2013, 3.

distribution programs is growing rapidly. We need a right to privacy in public to ensure that new data collection programs do not create personal or societal privacy harms.

## Privacy in Practice

Locational privacy is difficult to enact in a meaningful manner for many reasons, including the rapid expansion of the online environment, the complexity of federal and local laws that govern information distribution, the difficulty of providing effective notice to users, and the challenge of designing privacy policies in a way that gives users the opportunity to actively consent to or withdraw from a data collection program. This section of the literature review provides a brief overview of some of the laws that affect data collection programs in the United States and seeks to demonstrate why typical notice and consent policies do not adequately protect locational privacy.

### Laws governing privacy and transparency

The legal dimension of privacy is too large in both breadth and depth to be surveyed comprehensively in this literature review. Instead, I will draw on some key pieces of law and legal scholarship that have shaped the modern debate over privacy.

#### *Origins of Notice and Consent in the Law*

Interestingly, there is no federal law that requires the use of a notice and consent regime to protect personal privacy for most types of data collection programs.[40] Using notice and consent regimes to protect personally identifiable information (PII) began in 1973 with the adoption of the Fair Information Practice Principles (FIPP) by the U.S. Department of Health, Education, and Welfare (HEW). HEW was concerned about the

[40] Lane 2014, 65.

increasing rate of digitization of data and outlined five principles to help protect personal data from being used for new purposes without consent. The four principles were:

1) Transparency of record systems of personal data;
2) The right to notice about such record systems;
3) The right to prevent personal data from being used for new purposes without consent;
4) The right to correct or amend one's records;
5) The placement of responsibility for preventing data misuse on the data holders.[41]

These principles have shaped both privacy policy and privacy law in the last four decades and have led to the ubiquity of notice and consent privacy policies. FIPP is limited in the types of information that it protects; because it was designed before data aggregation and analytics were commonplace, it is only concerned with PII like name, social security number, financial information, and sensitive health information. It does not protect locational information like home addresses and has left many forms of online tracking information unprotected.[42] Because of this oversight, geospatial and spatio-temporal data are not protected forms of information, and it is possible within the scope of FIPP to collect other forms of identifying information at the individual level. The fact that FIPP was created in 1973 and has not been noticeably updated since 1996[43] means that countless forms of information and metadata are not protected by law.

FIPP is content-neutral; if a person is given notice of a policy and consents to it, FIPP has nothing to say about the relative merits of that information exchange. FIPP assumes that people are rational actors when they consent to share information; they

[41]Solove, Daniel. "Introduction: Privacy Self-Management and the Consent Dilemma." *Harvard Law Review* 126, no. 7 (2013): 1880, 1883.
[42] Nissenbaum 2010, 56,
[43] Finch and Tene 2014, 4.

ideally would evaluate the costs and benefits of any exchange of information and agree only to those in which there was a clear benefit to the exchange. Research in behavioral economics and psychology has demonstrated that this ideal of rational decision making is a myth; because people have limited physical and mental resources with which to achieve a wide variety of tasks and goals, people tend to make decisions using shortcuts or heuristics to quickly assess problems. Because notice and consent privacy policies tend to be long, legalistic documents that do not offer options for varied levels of consent to information sharing, it makes sense that people are unlikely to actively engage with notice and consent privacy policies.[44] Daniel Solove makes this point when he argues that privacy self-management, the idea that each individual is responsible for managing their personal information, is a flawed system. Asking people to develop heightened awareness of how data is used after it is released is the same as asking someone to track the precise dollar they used to buy a cup of coffee as it is used time and again in countless other transactions. Successfully managing personal privacy through the privacy self-management is not realistic given the complexity of the information economy. Thus, a regulatory agency is needed to protect privacy through the use of law and policy. In the last twenty years, the FTC has filled this role.

### *Federal Trade Commission as Protector of Online Privacy*

In the last two decades, the Federal Trade Commission (FTC) has emerged as the de facto regulator of online privacy and consumer protection. It has begun to enforce FIPP and has managed to restructure its regulatory powers to address a new era of consumer concerns.

[44] Solove 2013, 18.

The FTC is perhaps an unlikely protector of online privacy. The agency was created in the first decade of the twentieth century to act as an antitrust agency and as a regulator of false advertising and corporate espionage. Its legal authority is the Federal Trade Commission Act, which draws on elements of common law to give the agency the power to set forth regulatory policies on unfair practices and to evaluate companies on their adherence to self-imposed standards. In *Federal Trade Commission: Privacy Law and Policy*, Professor Chris Hoofnagle explains that part of the reason the FTC has remained a potent regulatory agency over the past century is that it has the power to regulate cases where common law principles have been violated.[45] By linking innovative (and perhaps unfair) trade practices to common law principles, the FTC has been able to transform from an antitrust and monopoly regulatory agency to an agency that is capable of regulating online privacy. However, the FTC does not have civil penalty power and instead relies on the fact that FTC regulatory action creates negative press and damages brand reputations for companies that violate consumer's expectations of privacy.[46] Even with this limitation, the FTC has become an agency that can ensure that companies comply with the privacy policies that they create and that data collection practices do not violate contextual norms.

Part of the FTC's success in establishing its ability to regulate online privacy comes from the fact that it is very careful in the type of cases that it chooses to pursue. Most of its cases have to do with protecting the privacy of children or preventing companies from retroactively changing their privacy policies in ways that renege previously-made

[45] Hoofnagle, Chris Jay. *Federal Trade Commission Privacy Law and Policy*. New York: Cambridge University Press, 2016, 30.

[46] Hoofnagle 2016, 146.

promises. The FTC brought its first major online privacy case in 1997, long before online privacy was a concern to the general public, and declared that the same privacy principles that apply in traditional marketing apply online, especially when it comes to fraudulent advertising and data collection practices targeted at children.[47] After hosting a conference on the future of online privacy, the FTC adopted four principles that have come to be the de facto norms of online privacy regulation:

1. Notice: Data collectors must disclose their information practices before collecting personal information from consumers.
2. Choice: Consumers must be given reasonable options with respect to whether and how personal information collected from them may be used for purposes beyond those for which the information was provided.
3. Access: Consumers should be able to view and contest the accuracy and completeness of data collected about them.
4. Security: Data collectors must take reasonable steps to ensure that information collected from consumers is accurate and secure from unauthorized use.[48]

Through application of these principles on a case-by-case basis, and by drawing on privacy norms already established in fields like health, domestic sanctity, and children's rights, the FTC became a norm entrepreneur in the 1990's and eventually created a standard of self-regulation that led to the widespread adoption of privacy policies and internal privacy assessments at all major internet companies.

However, now that the role of data brokers and shortcomings of privacy self-management through notice and consent have become clear, the FTC is shifting away from enforcing privacy representations as expressed in privacy policies to enforcing consumer expectations. This means that the FTC is now interested in ensuring that consumers online understand what companies do with data they collect and that the companies actually do as

[47] Hoofnagle 2016, 157.
[48] Hoofnagle 2016, 153.

they promise. One of the major impacts that this has had on the FTC is that the agency is now interested in promoting privacy by design (PBD), which is the idea that privacy can best be protected through proactive design of systems that avoid collecting and storing sensitive information. In the words of Professor Hoofnagle:

> PBD encourages system designers, who have previously thought that privacy issues were for policy makers to shape, to think through the ethical implications of their technologies from the beginning and to consider building privacy protections into products.[49]

Instead of focusing on retroactively protecting privacy, PBD embeds privacy principles into the hardware and software of new technologies. PBD can take many different forms based on the unique needs of each individual system, but building privacy into systems from the ground up makes it easier to manage data that is collected in a way that protects individual privacy. The complexities of PBD will be covered in greater depth in a later section, but the fact the FTC has embraced PBD shows that PBD has the potential to dramatically shift the nature of the online privacy debate.

While the FTC has become the de facto regulator of online privacy and has pushed companies towards adopting enforceable privacy policies, it is not designed to regulate government agencies. Because its authority is drawn from antitrust laws, it does not have the regulatory scope to determine what public entities do to protect privacy. From the perspective of looking at locational privacy concerns in municipal data collection programs, this is a large gap in the regulatory framework.

Also, the FTC has never taken a case specifically involving locational privacy. Instead, the FTC focuses mostly on cases where online companies have clearly violated their own

[49] Hoofnagle 2016, 178.

privacy policy in a way that violates consumer's expectations. The FTC has also been successful in regulating privacy when companies are specifically targeting children. Because the FTC does not have explicit authority to regulate privacy, it is likely that if it was going to regulate locational privacy, it would have to be somehow related to an issue of consumer expectation or children's rights.

### *Open Government Directive*

While the FTC works to protect privacy online, the White House Open Government Initiative has embraced the principle of government transparency by encouraging agencies at all levels of government to make as much data available to the public as possible. In 2009, the Obama Administration issued the Open Government Directive, which instructed all federal government agencies to take steps to ensure that agency information was available online in a reproducible, complete, machine-readable format. Since 2009, over 190,000 datasets have been published on data.gov, covering subjects from agriculture to urban crime rates.[50] While the Open Government Directive only legally applies to federal agencies, the Obama Administration has encouraged cities to upload datasets and has invested millions of dollars in smart city technologies, with the stipulation that the data generated by the program be made accessible to the public.[51] The push into data transparency at both the federal and local level has led cities large and small to publish datasets in open data portals in an attempt to both comply with the President's plan and to lower the cost of public records requests by proactively publishing datasets.[52]

[50] "Open Data Portal." Last modified March 18, 2016. Accessed March 22, 2016. http://www.data.gov/ ,2.
[51] Office of the Press Secretary 2016, 1.
[52] Whittington et al 2016, 1936.

### *Freedom of Information Act*

The Freedom of Information Act of 1974 (FOIA) is a federal information transparency law that allows for the full or partial disclosure of documents controlled by the United States federal government. Any person or corporation can make a FOIA request with any federal agency. Agencies are required to answer all requests either with the requested information or with an explanation of how the data requested violates one of the nine exemptions to the law (exemptions include national defense, trade secrets, and medical files). The law does not protect locational information in any way.[53] Requests for information under FOIA can be incredibly broad; in some cases, responding to FOIA requests can require hundreds of man-hours because of the breadth of the request. This is becoming increasingly true in the era of big data; collecting all pertinent records related to a topic can be a herculean task.

Every state has a local version of FOIA that allows residents, journalists, and corporations to request information created by state or city government agencies. These laws vary in scope; some allow people to ask for almost any information, while laws in other states are much more limited. While a full enumeration of each state's information transparency laws would be tedious, a discussion of Washington State's Public Record Act is included below because it is the legal framework in which all three case studies considered in this thesis must operate.

[53] U.S. Department of State. "The Freedom of Information Act." U.S. Department of State. Last modified January 11, 2016. Accessed March 22, 2016. https://foia.state.gov/Learn/FOIA.aspx, 1.

### *Washington State Public Records Act*

Washington State's Public Records Act (PRA) is an especially broad and expansive information transparency law.[54] PRA requires that all records maintained by state and local agencies be made available to all members of the public. The law applies to all information that is prepared, owned, used, or retained by any state or local government agency that relates to the work of any governmental body. The law also requires that no form of identifying information can be required to place a PRA request.

The exceptions to this law are few, and must be demonstrated any time a request is denied. Exemptions must be applied narrowly, and only cover limited categories of information including the identities of children or hospital patients, financial information, vehicle license plate information for publicly owned vehicles, the identities of sexual abuse victims, jail records, voter registration and voting histories.[55] While other exemptions exist, none would offer wide-scale protection for locational data. There is no exemption that allows for agencies to deny requests for information because the requested information includes unique identifiers that could be used to link records across separate datasets. For the most part, the exemptions that exist are based on unique characteristics of individuals rather than on the data type generally (e.g. someone who had been in a traffic accident would not merit the same level of protection as a rape victim). This means that there are very few exceptions for database-level exclusion from release (meaning that the entire database is protected rather than just certain records). Indeed, the only database-level

[54] MRSC of Washington. "Public Records Act." MRSC. Last modified June 1, 2015. Accessed April 23, 2016. http://mrsc.org/Home/Explore-Topics/Legal/Open-Government/Public-Records-Act.aspx, 1.

[55] MSRC of Washington 2015, 2.

exclusion included within the PRA itself is the exclusion for financial information, which protects all information involved in a financial transaction.[56] Since it is the whole-sale release of massive datasets that can be used to trace movement patterns that is a threat to locational privacy, the PRA opens a door for the forced release of locational data that could be used to locate and track individuals in deeply unacceptable ways. Thus, the data collected through smart city data collection programs must be evaluated not only for its usefulness but also for its potential to be used for purposes that it was not originally intended for. This is especially important for data that contains spatial identifiers because geospatial information can be used to join datasets, infer new information, and potentially violate the privacy of individuals or groups.

### *City of Seattle's Privacy Policy*

The City of Seattle is a thought leader in the field of privacy protection. In September of 2014, the Seattle Department of Information Technology (DoIT) launched the City of Seattle Privacy Initiative. After convening a stakeholder group of privacy scholars, community members, and government employees from 15 city departments, the Privacy Initiative designed a citywide Privacy Program to guide data collection and management practices in all areas of city government. In February of 2015, the Seattle City Council adopted the six Privacy Principles, which provide an ethical framework for developing department-level policies, standards, and practices for data collection and management. The six principles are:

1. Value of Privacy: Privacy Impact Assessments will be completed on all new data collection programs.

[56] MSRC of Washington 2015, 2.

2. Minimization: The City will work to only collect what data is necessary for service provision.
3. Notice: The City will work to provide notice about how personal data is used and will provide an opportunity to opt-out whenever possible.
4. Accountability: The City will comply with all federal and state laws regarding privacy.
5. Transparency: The City follows all federal and state laws regarding public disclosure requests. Third party contractors with access to personal information must comply with the City's privacy policy.
6. Accuracy: The City will work to correct inaccurate personal information when practical.

These principles represent the many months of work with stakeholders and incorporate some of the privacy best-practices suggested by privacy scholars and advocates who advised on the process. The six principles reveal that the City must balance its commitment to privacy with many other concerns, including legal compliance with disclosure laws, provision of city services, and the commitment to transparency and citizen oversight. This is most noticeable in the careful language used to delineate the extent of the City's commitment to the principles of notice and accuracy. Because of its other commitments and technical constraints, the City cannot unequivocally commit to providing notice or commit to using an opt-in data collection system rather than an opt-out method. While sensible, the privacy principles show that the City's commitment to privacy is constrained by a wider set of municipal concerns.

The City Council took further action in July of 2015 and adopted an updated privacy statement based on the six Privacy Principles. The policy provides direction to all City departments about privacy practices and data management. It further requires a Privacy Impact Assessment and a Privacy Threshold Analysis for any new data collection program. Together these checkpoints ensure that new data collection programs uphold the six principles and comply with federal and state laws. The policy statement directs all

departments to provide notice about the collection, use, and sharing of personal information at the point at which data is collected and to provide an option to opt-out whenever feasible.[57]

The Privacy Policy is one of the strongest and most comprehensive privacy statements to date because of the City's broad definition of personal information:

> Personal information is any information relating to an identified or identifiable individual. Examples of personal information include but are not limited to a person's name, home or email address, social security number, religion, political opinions, financial and health records, location and racial and ethnic origin.[58]

This definition includes location, email address, and race and ethnicity, which are not typically included in legal definitions of PII. The breadth of this definition extends privacy protection to types of information that are not protected by state or federal laws. For example, locational privacy is not an exemption to the PRA or the FOIA, nor is it regulated by the FTC. While Seattle must still comply with PRA and other information disclosure laws, the inclusion of location in Seattle's definition of information privacy sends a strong signal about the City's commitment to personal privacy.

### *ORCA Privacy Policy*

Sound Transit also has a privacy policy that outlines its information collection, processing and distribution policies.[59] It is freely available on the ORCA program website but is not publicized at bus stops or at transit stations where transit passes are sold. While

[57] City of Seattle. "City of Seattle Privacy Policy." Last modified April 2016. Accessed June 6, 2016. http://www.seattle.gov/Documents/Departments/InformationTechnology/privacy/PrivacyPolicyFINAL.pdf.
[58] City of Seattle. "Privacy Facts." Last modified March 2016. Accessed June 6, 2016. http://www.seattle.gov/Documents/Departments/InformationTechnology/privacy/PrivacyFactsLB.pdf.

[59] Sound Transit. "ORCA Privacy Statement." Last modified May 14, 2012. Accessed March 22, 2016. https://www.orcacard.biz/ERG-Seattle-Institution/jsp/static/ORCA%20Privacy%20Statement.pdf.

this may be a result of cost reduction efforts, it limits rider awareness of the privacy policy. The ORCA privacy policy states that personal information is collected when people purchase an ORCA card or add money to it. It warns employees that trip information from cards paid for by employers may be shared with employers. Importantly, it notes that most of the information on individual cards is not encrypted; the exception is date of birth and pass type expiration date. The privacy policy clearly lists different types of information collected at various stages of the program and is careful to note which information transfers are optional and which are not. It also states that Sound Transit intends to only use data for legitimate agency purposes and will not distribute information unless required by law or court order. The policy explicitly acknowledges that trip data could be linked to other independently-collected datasets if released, and disavows responsibility for privacy harms resulting from data linkages.

The ORCA privacy policy goes on to address data retention and disclosure. It states that data will be retained for as long as the agency considers useful in both aggregated and disaggregated forms. In a section titled "Public Records," the agency specifically demonstrates the enumerated exemptions from the Washington State Public Records Act that grant it immunity from wholesale disclosure requests. However, it does note that information regarding the use of transit passes may be released in aggregate under the law if the data does not include any identifying information. This is crucial because it sets the stage for ORCA data to be released using a tiered data release program, which will be discussed in later sections.[60]

[60] Sound Transit 2012, 8.

## Notice and Consent

Privacy scholars typically consider notice and consent frameworks deeply flawed because they do not offer users meaningful choice, are usually difficult to understand, and often do not clarify how information is handled once it is collected.[61] Privacy notices can take different shapes and use different channels of communication, ranging from a privacy policy document posted on a website to signs posted in public places to inform passersby about CCTV cameras.[62] Each data collection program should have a privacy policy that is tailored to the method of data collection. For example, an online survey could ask participants to select their data sharing preferences through a series of yes/no toggle switches covering a range of data sharing scenarios while a public space data collection program might rely on signage directing people to more robust information resources. Privacy protection policies must be designed to work with the architecture of the data collection platform. There is no one-size-fits-all solution.

Part of the difficulty of notice and consent frameworks is that these frameworks aim to achieve two separate tasks with one tool. That is, data collectors are attempting to both notify people about the data collection program and obtain their consent for data collection. For data collectors, privacy notices serve multiple purposes including compliance with legal and regulatory frameworks and building user trust[63]. However, these

---

[61] Calo, M. Ryan. "Against Notice Skepticism in Privacy (and Elsewhere)." *Notre Dame Law Review* 87, no. 3 (2012): 1027, 1033.

[62] Gutmann, Myron, P. Witkowski, Kristine Colyer, Corey O'Rourke, and JoAnne McNally. "Providing Spatial Data for Secondary Analysis: Issues and Current Practices Relating to Confidentiality." *Population Research and Policy Review* 27, no. 6 (2008): 639-65.

[63] Schaub, Florian, Rebecca Balebako, Adam Durity, and Lorrie Faith Cranor. "A Design Space for Effective Privacy Notices." *Symposium on Usable Privacy and Security*, July 22, 2015. Schulz, Jerry. "Information Technology in Local Government." ICMA. Last modified October 31, 2015.Accessed December 4, 2015, 6.

goals are not always complimentary. For example, the legal language required by the FTC is dense and difficult for the average reader to understand, so a privacy policy that is well-crafted from a legal standpoint may not be an effective tool for communicating data practices with users. The complex network of goals that notice and consent frameworks must meet to protect privacy often leads to dense contracts and a binary choice for users: either agree to the privacy policy or forego use of the service. A recent Pew Research Center survey found that only 18% of American internet users say that they read all privacy notices that they agree to.[64] It is estimated that the average American internet user would have to spend 244 hours per year reading privacy policies if they were to engage with every privacy policy they encounter online.[65]

With these constraints, it is obvious that notice and consent frameworks leave much to be desired as a mechanism for privacy protection. Indeed, Chris Hoofnagle notes that instead of providing adequate protection for privacy, the use of notice and consent frameworks actually has the effect of deterring substantive conversations about the normative aspects of data collection:

> A robust privacy debate should ask who needs our data and why, while proposing institutional arrangements for resisting the path offered by Silicon Valley. Instead of bickering over interpretations of Facebook's privacy policy as if it were the US Constitution, why not ask how our sense of who we are is shaped by algorithms, databases, and apps?[66]

Far from protecting privacy, the provision of vague and opaque privacy policies limits the ability of consumers to understand how their data is being used, let alone take action to

[64] Pew Research Center 2015, 7.
[65] McDonald, Aleecia M. and Lorrie Faith Cranor. "The Cost of Reading Privacy Policies." I/S: A Journal of Law and Policy for the Information Society 4 (2009): 543-897.
[66] Hoofnagle 2016, 173.

protect their information. Because of the limitations of the typical notice and consent framework and the fact that privacy policies can act as a smoke screen for organizations that are not interested in incurring transaction costs by imposing privacy-minded self-regulation schemes, scholars, planners, and privacy advocates need to consider new ways of protecting privacy that go beyond notice and consent.

However, it is important to note that most of the available academic literature regarding notice and consent only covers online platforms; it does not address data collection programs that use sensors in public spaces. Privacy protection for public space data collection programs will be addressed more fully in a later section. In order to fully understand the challenges of using notice and consent to protect the locational privacy of residents within the context of smart city sensor programs, we must first discuss the components of a notice and consent framework, including different privacy notice designs and consent structures.

## Notice

Notifying users about a system's data practices is supposed to enable them to make informed decisions about information privacy. Notice is a distinct concept from consent, even though it is often presented to users at the same point where they are asked to consent to the data collector's terms of use. In modern privacy policies, these two distinct aspects are paired together because notice must be provided before users can consent to a policy, and also because the integration of the two concepts closely models typical legal contracts.[67] However, research has shown that privacy policies are often ineffective

[67] Schaub et al 2015, 3.

because they are neither usable nor useful. If the only choice presented to the user is either to consent and use the service in question or to not consent and forego the use of the service, it makes sense that users interact with privacy policies as little as possible.[68]

The problem of presenting notice in a format that is usable and useful to the person who is making the decision to use a service will only increase as ubiquitous computing platforms begin to collect more detailed and granular data. The current approach of viewing the privacy policy as a legal contract, replete with the language of legal experts, is not suitable to ubiquitous computing platforms because there is no "gateway screen" on which to present a privacy policy.[69] Delivery of notice cannot be handled through a click-through contract because sensor networks do not use web-based applications to gather data. In most cases, if a sensor network is working correctly, data subjects will not know that they are being watched.[70] Thus, to protect privacy in the age of ubiquitous computing, we need to engineer systems that either deliver notice in novel ways or we need to design systems that are themselves privacy-aware.[71]

## Consent

### *Opt-in/Opt-out*

Although notice of data collection practices is essential for consent, notice alone does not allow for privacy self-management. For people to be able to manage their own privacy, they must have meaningful options regarding what information they are willing to

[68] Barocas and Nissenbaum 2015, 14.
[69] Schaub et al 2015, 6.
[70] Konings, Bastian, Schaub, Florian, Weber, Michael, Mattern, Friedemann, Santini, Silvia, Canny, John, F., Langheinrich, Marc, and Rekimoto, Jun. "Prifi Beacons: Piggybacking Privacy Implications on Wifi Beacons." *Pervasive and Ubiquitous Computing Adjunct Publication Proceedings of the 2013 ACM Conference*, 2013, 83-86.
[71] This will be covered in greater detail in the "privacy by design" section of this chapter.

share. Consent regimes are typically structured around whether users are given the option to "opt-in" to or "opt-out" of the data collection program. When data collection programs are structured around an "opt-in" framework, each user must give explicit consent to data collection before any data is collected. The most common structure for opt-in consent is a "click-through" terms of use, where users must actively engage with the privacy policy and agree to be bound by the terms of the contract before they can continue to the service that they are interested in. This is known as a binary opt-in structure because people must either agree to the policy or choose to forego use of the service.

In an ideal opt-in framework, individuals are given a high degree of control over what data is collected and how it is used. For example, a fitness tracking app could ask users for permission to share data with public health researchers. The app could offer levels of sharing; at the most permissive level, all information including name, age, and location could be shared. At a medium level of permissiveness, data would be aggregated and anonymized before it was released. If the user did not want to share any information, the data would not be distributed to public health researchers. While this is still not a perfect solution to privacy it is better than using a binary model where the only choice is either to use the service and share all data or forego use of the service entirely. Opt-in consent structures only provide user control over data if they offer a range of data collection and sharing choices. Thus, binary opt-in consent structures are not a strong foundation for consent-based privacy regimes.

Opt-out consent frameworks are also problematic. In a typical opt-out structure, data will be collected until a data subject takes active steps to remove themselves from the data pool. This is often used in situations where it is not possible to provide direct notice.

For example, some stores track the WiFi signals sent out by smartphones to track how many people visit the store, record which areas they visited, and determine whether or not they bought anything.[72] Because the data collection is tied to a space rather than to a specific webpage, it is difficult to identify the data subject and give them the opportunity to opt-in to data collection. Some potential forms this could take might be either a sign at the front of the store explaining the data collection program and providing instructions on how to opt-out of the data collection effort.[73]

However, this method is imperfect. To return to the example of the store, it is possible that people would not see the sign. Also, the sign distributes notice to everyone who enters the store rather than just notifying the people who have WiFi signaling activated on their smartphones. Thus, instead of using an opt-in consent structure, the store (or its software provider) hosts an "opt-out" form on its website for anyone who is uncomfortable with the program. This option is frequently used for space-based data collection programs because it is not intrusive and is targeted only at data subjects. Unfortunately, these opt-out consent options are rarely paired with a robust privacy notice; oftentimes, people are unaware that data is being collected and that they have the option to opt-out of the program. Obviously this presents a challenge for opt-out consent structures that must be rectified if privacy rights are to be adequately protected.

Essentially, whether or not a privacy policy is designed around opt-in or opt-out consent depends on the ease of delivering notice to data subjects. When it is relatively easy to present every data subject with a notice statement and ask them to agree to it, it is likely

[72] Lane 2014, 84.
[73] Schaub et al 2015, 5.

that the program will use an opt-in framework. In situations where presenting notice is not so easy or data collectors cannot be sure that everyone will be presented with notice in the same way, it is likely that they will use an opt-out format. However, it is important to recognize that the fact that providing notice and consent is not easy should not be a justification for unchecked data collection. Being informed of data collection and given a voice in how it is used promotes individual engagement with data collection practices and restores a level of control over data distribution to the data subject. Even if notice and consent structures are not easy to implement or design in a way that is truly protective of privacy, program managers need to find a way to restore a degree of choice and awareness to data subjects.

### Granular consent

One of the major problems with opt in/opt out consent structures is that they offer a binary choice: either the data subject agrees to the privacy policy in full or they choose to not use the service or space where data collection is taking place. This may be an acceptable framework in situations where people have other choices for obtaining the goods and services they need, but in the context of municipal agencies, people may be required to interact with the city. If interaction with the city requires consenting to the privacy policy, people may be de facto *forced* to participate in the city's data collection program. This form of consent is at best compelled consent and can be contrasted with free consent, the condition that exists when potential data subjects can satisfy their needs via other providers without consenting to data collection practices they find objectionable. The problem of compelled consent affects all aspects of city government from building

permitting to the fire and police departments. If people *must* use city services, the binary consent structure of both opt in and opt out frameworks breaks down and can no longer honestly be deemed free consent.

One of the best ways to mitigate the tension between compelled use of a service and free consent to data collection is to provide granular control over which data collection initiatives a data subject actively consents to. This consent structure is still being developed[74][75], but one way this framework could be implemented is through "as-needed" consent. In this framework, people are asked for their consent as they progress through the various aspects of the service. If they choose to not consent to data collection, the logic of the software platform excludes the data associated with their persistent identifier from further data collection in accordance with the privacy preferences they have declared. Even if a person chooses to not participate in a certain aspect of the data collection program, they are still able to use the service as they see fit.

As-needed consent frameworks are well-suited to use in municipal data collection programs because they preserve the freedom of choice on which consent is based. In other words, as-needed consent frameworks give people some degree of control over what information is collected and what can be done with their information. For example, when people sign up for a new public utilities account, the account setup process could describe the data collected by this technology and include a series of questions about participation

---

74 Information and Privacy Commissioner of Ontario. Transparency, Privacy, and the Internet: A Municipal Balancing Act. Ontario: City of Ontario, 2014, 5.

75 One potential challenge to granular consent is that if the shifting nature of informational norms are taken seriously, the type of choices that need to be presented to users would also have to evolve over time. This time-based fluidity is one of the major reasons that informational norms are not actively used in privacy aware systems.

in research on electricity usage. The utility could make it clear that anonymized metering data might be shared with third party researchers and that the data could be used to make service provision more cost-effective. Obviously, even if the new user chose to not participate in the data collection, they would still be able to receive utility services. By asking explicit questions about data sharing, cities can build a relationship of trust with residents. Residents who believe that the city respects their privacy will be more likely to share information and participate in voluntary data collection programs, which will make it easier to get robust data for research in the long run.[76]

In summary, commonly used consent structures leave much to be desired. Opt-in and opt-out consent frameworks are best seen as legal tools designed to meet the regulatory standards imposed by the FTC and state-level privacy protection agencies; they are not primarily designed to give individuals full control over what data is collected or how it can be used. Renewed concern about large scale data collection has led to new interest in granular consent frameworks like as-needed consent, but much research remains to be done to create consent structures that provide better control over data collection, use, and distribution.

## Challenges to Notice and Consent

As the online environment becomes more complex, the challenges of using notice and consent regimes to effectively protect privacy also grow. There are three looming challenges to the use of notice and consent regimes to protect information: 1) multi-source

76 Under RCW 42.56.330, records generated by a public utility are exempt from access by law enforcement authorities unless the law enforcement authority presents a written statement that the person the requested records are linked to is suspected of a crime.

data aggregation, 2) unforeseen data use, and 3) non-traditional data collection methods. Multi-source data aggregation, a now-common practice where analysts use a common attribute or logical relationship to combine records, is a challenge for privacy protection through notice and consent because it is almost impossible for data subjects to track *which* pieces of information they have willingly disclosed. It is impossible for someone to know which pieces of information will be combined in the future, or if seemingly innocuous pieces of information could potentially be combined to infer private information.[77] There are many types of aggregation techniques; for example, recent advances in computer science have demonstrated that it is possible to uniquely identify an individual with as few as four spatio-temporal data points in a geographic area as large as a city.[78] Given that detailed information about consumers is (and will continue to be) a valuable asset to businesses, it is likely that data analysts will continue to search for new ways to aggregate data about individuals. Attempts to protect personal information against aggregation by depending on individuals to manage what information they disclose to various data collection programs may prove to be futile.

Unforeseen use of data is also a problem for privacy protection through notice and consent regimes because it is impossible for data subjects to fully comprehend what will be done with their data in the future. Data is non-rivalrous, meaning that it is possible to copy and share a dataset without affecting the original dataset. This means that consent to data collection may result in much wider dissemination of information than originally intended by the data subject. While data sharing practices can be disclosed in the privacy notice, the

[77] Barocas and Nissenbaum 2014, 3.
[78] De Montjove 2013, 4.

clauses covering downstream data use tend to be vague. From the point of view of the data collector, vague language helps avoid questions of whether or not the data subject's consent will apply to future uses of the data. Obtaining consent for every reuse of the data may be prohibitively costly and would likely lead to the cessation of the analysis project.[79] However, from the point of view of the data subject, vague language regarding future use of the data makes it almost impossible to assess what will be done with the data. How can a data subject truly consent if she is not given enough information to know what it is she is consenting to? A well-designed notification regime can help limit the potential for unethical downstream uses of data. This is one of the strongest reasons to have a notification regime; it creates an ethical yardstick by which downstream data uses can be measured.

The final obstacle to using notice and consent policies to protect privacy is the growing number of data collection platforms that are not conducive to the use of textual notice and consent policies. Usually, a notice and consent regime asks a user to read a policy statement and agree to the terms outlined therein. This format works to some extent for situations like visiting the doctor or choosing to use an online social media platform, but many data collection platforms struggle to provide either notice or consent in a way that can help potential data subjects make decisions. Programs such as Automated License

79 Of course, there is a long-standing debate about whether or not data should be collected to answer a defined research question or if data should be collected and stored en masse in order to be able to answer future questions or solve future problems. This is the essence of the divide between the traditional academy, which tends to collect only data relevant to the research question at hand, and the emerging fields of data science and data analytics. As the price of information storage has fallen, it has become increasingly cost-effective to collect and store information. The question, then, is whether or not we should avail ourselves of this ability. The fundamental difficulty of answering this question is that it is difficult to weigh the potential benefit of data collection against the potential privacy violation that data reuse could pose. Since both quantities are unknown, it is difficult to take a well-reasoned stand on the issue. A full evaluation of the issue is beyond the scope of this literature review, but acknowledging the tension is an important piece of this research.

Plate Readers (ALPR), WiFi MAC address sensors, or any smart city sensor program that gathers data about individuals all face the twin problems of both providing notice and giving people a chance to choose whether or not to participate. Currently, many data collection platforms that use in-situ sensors use an opt-out structure where users can go to a website and enter their personal identifier (e.g. a MAC address) into a "do-not-notice" database that is used to limit future data collection.[80] A "do-not-notice" database is a record of people who have asked to be excluded from the data collection. When the sensor recognizes a MAC address on the "do not notice" list, it removes their location trace from the dataset. However, this structure does not apply retroactively and places a high burden on the user.[81] Data subjects are often entirely unaware of the data collection programs, and even after they learn of it, opting out requires having strong enough technical research skills to locate the opt-out mechanism and find a device's MAC address. For average citizens, especially people who are not digital natives, this is likely to be a difficult task.

To be clear, the argument here is not that public realm data collection programs that use in-situ sensors are trying to be disingenuous. The difficulty of providing notice and consent for data collection programs that rely mostly on sensors is well-documented.[82] For example, the City of Seattle's privacy policy reflects the difficulty: "While it may limit the services we are able provide, *where it is possible* we will present information about what we are collecting and provide an opportunity to accept or decline to provide it to us"

[80] Schaub et al 2015, 22.
[81] Data Science Studies meeting, University of Washington, March 17, 2016.
[82] Schaub et al 2015, 27.

(emphasis added).[83] The fact that the City felt that it was necessary to include the phrase "where it is possible" demonstrates that the City is aware that delivering notice and obtaining consent in a timely manner may prove to be impossible given the nature of certain data collection programs.

What then, is to be done? If notice and consent regimes cannot protect information privacy, should we give up on the idea of privacy entirely? Ideas for changing the structure of privacy protection abound. Daniel Solove has suggested that the United States government create an Information Protection Bureau that would protect people's privacy interests in the same way that the Consumer Protection Bureau protects people from fraudulent transactions.[84] Solon Barocas and Helen Nissenbaum argue in favor of a stronger culture of data ethics for data analysts and researchers, similar to the Hippocratic Oath that all physicians must take.[85] Kelsey Finch supports a privacy "one-stop shop" where residents could select their privacy preferences for all data collection programs operated by the city. Still other researchers have taken active steps to build privacy into the design of the hardware and software that collect data in public spaces.[86]

All of these approaches are legitimate attempts to remediate the problems presented by notice and consent regimes for technologies of surveillance. Each proposed solution focuses on different aspects of the complex problem of protecting privacy in the digital age, and each offers different benefits. While it would be nice to find a comprehensive

[83] Department of Information Technology. "City of Seattle Privacy Statement." City of Seattle. Last modified December 2014. Accessed December 4, 2015. http://www.seattle.gov/information-technology/privacy-program/privacy-statement, 1.
[84] Solove 2013, 19.
[85] Barocas and Nissenbaum 2014, 5.
[86] Carnegie Mellon University 2016 and Schaub et al 2015, 15.

solution for privacy protection, both the complexity of the issue and the rate of change in the Information and Communication Technology (ICT) field are great enough to defy simple solutions. To better understand the opportunities and constraints of each proposed privacy protection method, it is important to stop and evaluate each platform in turn. Even though each proposal would need to be versatile enough to protect privacy in a variety of information mediums if it were to be used as a stand-alone policy, for the purposes of this literature review, I will tailor the discussion to how these proposed privacy protection methods can be used in public space data collection programs.

## Information Protection Bureau

Daniel Solove, a leading scholar of privacy law, argues convincingly that privacy self-management (the use of notice and consent regimes to give people the ability to "choose" what data sharing to engage in) is both necessary and flawed.[87] On one hand, current notice and consent regimes do not promote active engagement with privacy policies, but on the other hand, requiring notice and consent makes companies and government agencies more aware of privacy concerns regarding data collection and also gives data subjects some degree of agency in deciding what to consent to. As Solove points out, the ability to consent to activities that might be found questionable by society is one of the hallmarks of individual freedom.[88]

With that being said, Solove does believe that privacy self-management is not a stand-alone solution to personal privacy protection. He argues in favor of partial privacy self-management, a hybrid of notice and consent with a high-level regulatory agency that

[87] Solove 2013, 6.
[88] Solove 2013, 8.

would restrict certain pernicious data collection practices and provide people with a degree of certainty that their privacy was being respected while still giving individual end-users a chance to choose what level of information sharing they are comfortable with. He writes:

> In essence, what many people want when it comes to privacy is for their data to be collected, used, and disclosed in ways that benefit them or that benefit society without harming them individually. If people have an objection to certain uses of data, they want a right to say no. But many people do not want to micromanage their privacy. They want to know that someone is looking out for their privacy and that they will be protected from harmful uses.
> …We trust that [the products we buy] will fall within certain reasonable parameters of safety. We do not have to become experts on cars or milk, and people do not necessarily want to become experts on privacy either….People want some privacy self-management, just not too much.[89]

Solove is seeking a solution that strikes the balance between an overly rigid privacy protection agency that makes data collection prohibitively expensive and the weak form of "anything goes" consent we currently employ. By setting up parameters for acceptable information collection, the privacy protection agency could help to protect data subjects from revealing information that they may not want to give away or may not even be aware that they are giving away. Solove acknowledges that implementing a single agency to manage information privacy across millions of actors, information flows, and data collection platforms will be a challenge. Even with the difficulties of implementation, the idea of the government taking active steps to preserve privacy across the board is a logical solution to the problem. Because personal privacy is widely accepted as a foundational part of a liberal democracy with positive externalities both for individuals and the society as a whole, it is the responsibility of the government to correct the market failure of privacy violation through regulation and careful privacy management.

[89] Solove 2013, 12.

Right now, the FTC is filling this regulatory role. As discussed above, the FTC has molded itself into a regulator of online privacy by relying on antitrust and consumer protection regulatory power. It is not specifically tasked with protecting online privacy, so it lacks the ability to regulate in a comprehensive way and instead chooses to create rules through case law and case-by-base regulatory action.[90] This might not be the level of regulation that Solove would like to see, but it sets the stage for more comprehensive privacy regulation in the future.

Solove's solution of a high-level privacy protection agency that would moderate what types of data collection people could consent to is attractive for locational privacy protection because locational data is usually gathered through platforms that have no obvious notice framework. If people are unaware that their movements are being tracked by RFID tags, WiFi sniffers, or automatic vehicle license plate readers, it is impossible for them to manage their own privacy through a typical consent process. Ideally, the privacy protection agency would have a high standard for data collection platforms that were not able to deliver consent before collecting data; perhaps the agency would require a data management plan and would conduct periodic audits of the system to ensure compliance. While it would clearly be important that the agency not overly limit innovation or constrict the freedom to consent, a single agency that could review both data collection and downstream uses of the data would be greatly preferable to the current system of notice and consent.

[90] Hoofnagle 2016, 175.

## New Standards of Big Data Research Ethics

Helen Nissenbaum has been theorizing about informational privacy for over twenty years; she has reshaped the debate over information privacy through her work. She has developed many plans to manage privacy risk over the course of her career, but in conjunction with Solon Barocas, she has come to the conclusion that big data analytics have progressed to the point that one individual's consent or lack of consent has no bearing on what information is either directly collected (if the data subject consents) or inferred from statistical cross-section analysis (if the data subject does not consent). Barocas and Nissenbaum note that, "Inferences can be made about an entire population even if only a small fraction of people who share no ties are willing to disclose….Multiple attributes can be inferred globally when as few as 20% of the users reveal their attribute information".[91] Thus, even people who choose to opt-out of a data collection program can be swept up in the tidal wave of big data.

Barocas and Nissenbaum argue that notice and consent policies cannot be the sole gatekeepers of informational privacy; the relationships between actors, types of information collection, and potential downstream uses of data have all become too complex for even sophisticated users to navigate successfully. They also point out that notice and consent regimes originated in the field of medicine, which has a rich ecosystem of ethics already in place to ensure that physicians do no harm. Instead of relying solely on notice and consent to protect individuals from poor experimental design or risky procedures, medical professionals use an Internal Review Board to vet proposed research

[91] Barocas and Nissenbaum 2015, 19.

designs well before research subjects are asked for their consent. By ensuring that research proposals are scrutinized by experts, the medical field as a whole works to protect patients. In the medical field, asking for consent is a necessary but not sufficient condition. That is, notice and consent regimes are not enough to protect both the physical safety and the informational privacy of patients.[92]

Because FIPP and the original information privacy protection laws have roots in the medical field, it makes sense to look to the medical field for ways to protect personal privacy in the era of big data. Nissenbaum argues that notice and consent regimes are fundamentally flawed; for all the reasons discussed above, there is no way that a notice and consent regime can ever truly be robust enough to protect privacy. Instead, we need an ethical review board of experts that can evaluate proposed data collection, use, and distribution plans. The implementation of a data Internal Review Board (IRB) would promote better industry ethics and assure people that their data was being used in ethically responsible ways.[93]

Although this plan seems like a sensible proposal, many hurdles stand in the way of the use of IRBs to curtail violations of privacy. First off, data analytics is a toolset, not a unified profession like medicine or law. Anyone with the skills and access to datasets can derive new information. The fact that data analytics is not a unified profession makes it difficult to enact a set of professional ethics like those that govern doctors. Medical IRB relies both on doctors' internal ethics and on the threat of losing a medical license if the ethical violation is bad enough. While data analysts may have ethical moorings, there is no

[92] Barocas and Nissenbaum 2014, 2.
[93] Barocas and Nissenbaum 2014, 3.

such thing as a "data analyst license," so the data IRB would not be able to prevent unethical analysts from continuing in their objectionable practices. Additionally, the main benefit of big data analytics is that it has the power to derive new information from previously opaque datasets. This ability to reconstruct information, to target specific subpopulations, is a major source of the appeal of data analytics for firms. If compliance with the data IRB compromised the analyst's ability to derive useful information from their datasets, they would have vested interest in *not* complying with the injunctions of the IRB. If the data IRB is to succeed, there must be a clear set of incentives for firms and data analysts to comply with the ethical standards set forth by the data IRB. The only way the new data IRB would have enough sway over the industry would be if it was entrusted with governmental regulatory powers by both federal and state lawmakers.

While implementing new laws to control private sector use of data analytics may be a tall order, it may be possible for government agencies to lead on implementing a data IRB for in-house data collection efforts. City agencies are unique in that part of their explicit motivation is to protect the public good; they are not motivated by profit to the extent that a private firm must be. Thus, city agencies have the freedom and the responsibility to hold themselves to a higher ethical standard when it comes to data collection and management. This is an important first step because city agencies collect sensitive data about residents. If the city chooses to collect less data, or aggregates records before releasing them, it may be possible to at least make it harder for unethical actors to assemble data profiles about residents. At the very least, the city will be making a clear point about the value it places on residents' privacy.

A change in city policy can help to mitigate the effects of big data analytics on personal privacy, but the fact remains that until federal and state law adapt to the changed information environment, privacy protection will continue to suffer from an incomplete patchwork of policies and frameworks. The fact that the FTC currently only requires notice and consent for PII, and does not protect many important types of information like metadata and geospatial information is a major challenge to locational privacy. Given that public records laws like the PRA require that government agencies divulge any record upon request unless there is a specific exemption, it is important that state and federal laws be amended to protect privacy more fully. In the age of big data and ubiquitous computing, relics like notice and consent in simple opt-in or opt-out frameworks are no longer enough to protect privacy. There needs to be a sea change of legal scholarship and law-making to protect privacy in the modern age.

## Transparency in Pursuit of Privacy

In "Welcome to the Metropticon," Kelsey Finch and Omer Tene write about the perils and promises of smart city infrastructure systems.[94] They are optimistic that smart city infrastructure systems like smart grids, apps to report pot holes, and Intelligent Transportation Systems (ITS) can be used to provide better city services at a lower cost.[95] However, they also point out that cities that want to adopt smart city infrastructure need to be careful to ensure that new technologies benefit all residents equally and that the new programs are transparent.[96] They warn that without safeguards for transparency and

[94] Finch and Tene 2014, 2.
[95] Finch and Tene 2014, 25.
[96] Finch and Tene 2014, 32.

equality, smart city programs can become Orwellian nightmares[97] where sensors monitor every aspect of life from garbage disposal to travel patterns. For Finch and Tene, transparency about data collection ensures that residents will have a chance to speak up in favor of privacy. For Finch and Tene, the solution to the privacy challenge is to make the process of data collection more transparent. Ideally, the public would have access to as much detailed information as they want about the data collection technology. Where this is not possible for security reasons, the city would need to have a mechanism for answering questions on a personal level, as well as a high-level privacy policy that controls all forms of data collection in the city.

Finch and Tene also advocate for a mechanism similar to the Obama administration's Green Button program that would give residents the ability to access the city's records on their own resource consumption and travel patterns.[98] For example, a city with a smart grid could provide detailed by-appliance electricity consumption charts to each household to help residents reduce unnecessary use of electricity. This granular form of demand management has proven to be successful in test programs and could have the added side benefit of making people feel more comfortable with the city's use of smart grid technology.[99] The same platforms could also be used to provide notice and obtain granular consent to proposed uses and users of data. If implemented successfully, it is possible that these transparency measures would in fact shift the contextual norms related

[97] For more, see the podcast "Welcome to Nightvale".
[98] Finch and Tene 2014, 34.
[99] Finch and Tene 2014, 35.

to smart city sensor programs to the point that the public would not see the use of smart city infrastructure as an invasion of personal privacy.

While transparency is important, and it is essential that residents feel comfortable with the city's collection, use, and distribution of data, it is equally crucial that data is *actually* protected from pernicious use. A dashboard of personalized statistics might help individuals feel more comfortable with the data collection programs, but it will not stop companies from using public records requests to obtain large datasets that can be combined with other data sources to infer new relationships about individuals and groups. In a way, public engagement tactics that stress the benefits of smart city data programs while minimizing discussion of the risk have the potential to be disingenuous. If residents are going to be active participants in the process of deciding what smart city programs they are comfortable with, it is important that they be apprised of both the promises and perils of the smart city. Thus, it is not enough to inform residents of data collection efforts; experts need to work to ensure that risks are minimized and controlled even when the risk is not apparent to the public.

## Privacy by Design

Privacy by design advocates believe that the best way to protect privacy is to use the very technology that creates a problem to solve it. Privacy by design has seven foundational principles which, when followed, protect privacy across multiple technological platforms. The principles are:

1. Proactive design, not reactive solutions.
2. Privacy should be the default setting.
3. Privacy should be embedded into design.
4. Full functionality—Privacy-aware systems can still collect high-quality data.

5. End-to-end security—Full lifecycle protection.
6. Visibility and transparency—Privacy does not have to limit openness.
7. Respect for user privacy—Privacy regimes should be user-centric.[100]

This approach, which proactively works to protect user privacy while still allowing for the provision of high-quality service, has gained support around the world. For example privacy by design was unanimously passed as an International Resolution by the global assembly of Privacy Commissioners and Data Protection Regulators in Jerusalem in 2010.[101]

Privacy by design can apply to all types of data collection technologies and can take many different forms accordingly, but for the sake of this literature review, I will focus on privacy by design technologies that could be used to protect locational privacy. In general, there are two ways to use technology to protect locational privacy. One way to protect locational privacy would be to design data collection systems that avoid collecting or storing disaggregated data that could be associated with an individual. This approach is often referred to as localized aggregation technology. Another way to reach the same goal would be to create software that could communicate global privacy preferences both to sensors and to online applications, which is known as a privacy preference communication technology. This next section will discuss the possibilities and limitations of both approaches.

[100] Sempra Energy Utility. "Applying privacy by design." Last modified June 2012. Accessed March 22, 2016. http://www.sdge.com/privacypaper, 24.
[101] Sempra Energy Utility 2012, 34.

## Localized Aggregation Technology

Localized aggregation of data helps to reduce the overall computational load of a sensor network, makes the network more energy efficient, and may be used to protect locational privacy. Aggregation algorithms take individual data points and reduce them to summary statistics based on parameters established by system designers.[102] Usually, aggregation occurs at the sensor level before any data is sent to a centralized server or database. The precise methods of aggregation depend on the design of the sensor network and on what data the system designer wishes to retrieve from the network.[103] While using aggregation to protect locational privacy by creating summary statistics of individual locational data points might lead to some loss of research potential from the dataset, the ability to build locational privacy into sensor networks would be the simplest way to protect residents' locational privacy. By using an aggregation algorithm in the design of the sensor network, the city could ensure that the data would not be used to track individuals as they move through the city because the disaggregated data would not be sent to a central server. Nodal aggregation statistics should, in most cases, be sufficient for city agencies to make real-time decisions.[104] This type of privacy protection is ideal because under this system, data on individuals cannot be requested through a records request (e.g. a PRA request in Washington State) nor can it be stolen through a cyber-attack. After all, data that doesn't exist can't be stolen.[105]

---

[102] Rajagopalan, Ramesh and Varshney, Pramod K., "Data aggregation techniques in sensor networks: A survey" (2006). Electrical Engineering and Computer Science. Paper 22, 15.
[103] Rajagopalan and Varshney 2006, 16.
[104] Sempra Energy Utility 2012, 4.
[105]Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. Ethics and Information Technology, 12(4), 313-325.

## Privacy Preference Communication Technologies

One of the most promising potentialities of privacy by design is the idea that interactive ICT systems can be designed to communicate privacy preferences as metadata collected with every sensing transaction. Every time a WiFi sniffer, smart grid inverter, or camera records information about an individual, it would also record their pre-defined privacy preferences and handle information accordingly[106]. A privacy-aware sensor system could also broadcast its presence to personal devices, triggering either a machine-to-machine communication or a presentation of a more traditional statement of notice to the data subject. Research has been done to demonstrate that automated privacy preference disclosure is possible for WiFi networks, Bluetooth devices, and cameras. Each system uses already-established channels of communication to "piggy-back" privacy information into the data exchange.[107] As with most information flows, machine-to-machine privacy communications can flow both ways: sensor networks can broadcast their presence to the devices that they are tracking and generate notice statements for potential data subjects, and personal devices can broadcast privacy preferences to sensor networks in their vicinity. Both types of automated privacy protection have shown promise in research trials.

The interactive communication of privacy preferences from one computer to another has the ability to make privacy much easier to protect. Sensor networks broadcasting their presence and data collection practices has been demonstrated via WiFi networks[108] and has been used in conjunction with Android OS devices, webcams, and

[106] Konings et al 2013, 85.
[107] Konings et al 2013, 83.
[108] Konings et al 2013, 84.

even Roombas. The use of a privacy beacon is a fast, reliable and simple way to announce privacy implications of a sensor system in a user-friendly and transparent way. While this method does not directly allow for a data subject to opt in or out of data collection at the time of data collection, it could be paired with a consent mechanism such as a website that could offer data sharing options for those who were concerned with the data collection program.[109]

Data subjects' personal devices can also be used to automate privacy protection. With the right software, people could select one set of global privacy preferences for each of their electronic devices. The software would then send out a set of privacy preferences to any nearby sensors using either a WiFi signal or some other sort of signaling mechanism as appropriate.[110] This method allows people to have control over what information is shared, based on their understanding of the applicable informational norms, while still avoiding the many pitfalls of notice and consent regimes. This model may allow more choice than aggregation techniques because it allows every individual to choose their level of data sharing, similar to an opt-out consent structure wh. In an aggregation model, data is aggregated based on attributes about the data rather than user preference. However, is the software-based approach is more difficult to implement because it requires every user to have the correct software on their devices and it requires all sensor networks to use the same set of code to interpolate privacy preferences.[111] This is not an impossible task, but it

[109] Konings et al 2013, 84.
[110] Ashok, Ashwin, Viet Nguyen, Marco Gruteser, Narayan Mandayam, Wenjia Yuan, and Kristin Dana. "Do Not Share! Invisible Light Beacons for Signaling Preferences to Privacy-Respecting Cameras." Semantic Scholar. Last modified September 7, 2014. Accessed March 22, 2016. https://pdfs.semanticscholar.org/9f8d/cd4a1f08c430702c05e0c49a7ad411cba38c.pdf, 1.
[111] Konings et al 2013, 3.

requires a much higher degree of user awareness than the sensor-level privacy announcement discussed above.

Contextual integrity can be used to make the idea of machine-readable privacy preferences applicable to many different data collection platforms. One of the challenges of using a set of global privacy preferences to control every information exchange a cellphone, car, or computer participates in is that the norms of information distribution and flow are likely to vary widely depending on context. If globalized privacy preferences are to work effectively, they must be both logical and provide nuanced control over the types of (dis)allowed information transactions. Contextual integrity can help because it is flexible enough to accommodate many different informational contexts and because it has already been converted into machine-readable format. Having a full ethical code of privacy available to use in machine-to-machine privacy preference communication ensures that the choices granted to users can be made to be robust enough to cover most, if not all, information exchanges.[112] In "Privacy and Contextual Integrity: Framework and Applications," Barth et al suggest using First Order Temporal Logic (FOTL), which is a way to code time-dependent roles of actors into a machine-readable format, to express norms by describing the actors that participate in each context, the roles that these actors play, and their knowledge states, all at a specific point in time. However, it is important to note that while it is possible to encode relationships related to contextual integrity, fully

[112] Barth, Datta, Mitchell, and Nissenbaum. "Privacy and Contextual Integrity: Framework and Applications." *Security and Privacy, 2006 IEEE Symposium on*, 2006, 15 Pp.-98.

enumerating every possible relationship between actors would place a large computational burden on the application and would be likely to result in reduced functionality.[113]

One possible way to resolve this difficulty is to find a set of informational norms that apply to categories of actors, roles, and knowledge states. This avoids the problem of infinite recursion encountered in Barth's attempt to formally express all potentials of contextual integrity while still allowing for some level of individual control over information exchanges. One way to begin to develop a non-arbitrary set of explicit informational norms is to crowdsource the work. This is exactly the approach used by Shvartzshnaider et al in "Crowdsourcing Verifiable Contextual Integrity Norms".[114] The team developed an algorithm to ask targeted questions about actors, roles, relationships, and the appropriateness of information flows to determine what level of information sharing was deemed appropriate by a wide cross-section of people. The crowdsourcing experiment showed that the use of crowdsourced norms could be help to build a context-specific framework of privacy rules. While this research is still in an early phase, there is much potential for the use of crowdsourced norms to define norms of contextual integrity on periodic intervals.

While it may appear that these approaches to privacy are overly technical, they can actually be used to great effect to ensure that cities find the right balance between data collection and privacy protection. Crowdsourcing norms is already part of a typical planning process; it is the digital equivalent of a public outreach and engagement

[113] Shvartzshnaider, Yan, Schrasing Tong, Thomas Wies, Paula Kift, Helen Nissenbaum, Lakshminarayanan Subramanian, and Prateek Mittal. "Crowdsourcing Verifiable Contextual Integrity Norms." 2016, 3.

[114] Shvartzshnaider et all 2016, 1.

processes. By giving the public an active role in designing data collection systems and incorporating their needs and desires into the system design, it is possible to both collect data and respect the privacy concerns of data subjects. By using the crowdsourced norms model discussed above, cities can tailor privacy policies to fit their residents. Residents will be able to decide what level of data collection they are comfortable with and, through a majoritarian process, will be able to have direct control over the level of data that the city collects. By incorporating the public in the process of norm creation and by basing privacy policies on contextual integrity, cities can provide a meaningful way for residents to "consent" to data collection that goes far beyond the typical notice and consent architecture.

## Connecting Research to Practice

The scholarship of information privacy is a diverse and quickly-growing field. Scholars across the world are working to both justify a robust right to privacy in the information age and to design policies and systems that can protect privacy. International conferences on privacy are a yearly occurrence, and the growing body of scholarly research on privacy demonstrates how crucial this issue is to both individuals and to society. However, the true value of this work is in the ways that it is (or is not) used to inform the creation and management of smart city sensor networks. While a systematic review of every design choice or policy relating to smart city sensor networks is well outside the scope of this paper, a general sense of how the theoretical literature relates to real-world practices can be obtained through focusing on three case studies. These same case studies will be the focus of my own work in later chapters, but in this section, I will

review the extant literature on the policies and system designs of the following three case studies.

### *Case Study #1: ORCA Card Data (Seattle, WA)*

The ORCA Card is the electronic transit pass used in the greater Puget Sound region by all transit agencies that are part of the Sound Transit consortium. It operates on both buses and light rail and is a cornerstone of inter-agency transit collaboration. The electronic transit pass uses a 13.56 MHz passive RFID tag that complies with Near Field Communication (NFC) specifications to communicate with stations on buses or at stations.[115] The pass uses kilobits of memory to store a unique ID number, the amount of money stored for fare payments, records of the last ten trips made, the valuation of the last five value-add transactions, and pass type information (e.g. reduced fare, employee pass, or student pass). If the pass is a reduced-fare pass for a senior citizen, the pass will also store the date of birth of the user. While the cards do not store personal information like name or address, ORCA data does raise concerns about locational privacy because the card's persistent identifier can be used to link trips over time, and the passenger type information can be used to selectively identify classes of riders as well. Programs like the reduced-fare pass for seniors, the UPASS program at the University of Washington, and the ORCA LIFT reduced fare pass for low-income riders all are pass types that identify very select sub-groups of the population. This type of information makes it easier to uniquely identify individual users. Travel data from individuals is also available to employers who purchase

[115] Swedburg, Claire. "ORCA Puts Ferries, Buses and Trains on One Ticket." RFID Journal. Last modified October 22, 2009. Accessed March 22, 2016. http://www.rfidjournal.com/articles/view?5320, 1.

passes for their employees.[116] Employer information can also help to uniquely identify individuals because the employer identification number is also stored on each ORCA card bought for the company's employees. With some reverse engineering, the card could reveal where a person works, the time and location of their last ten trips on transit, how much the rider paid for their transit pass, and the rider's date of birth.[117]

While this is troubling to some degree, disaggregated transit data would be difficult to use to track individuals on a wide scale. Instead, the larger threat comes from the database of ORCA transactions, which records every trip made since the implementation of the system. This database is held by Vix, an Australian software company that contracted with Sound Transit to manage the data created through the ORCA program. The specifics of the ways in which Vix manages ORCA data, including whether or not it is allowed to sell ORCA information to other parties, is unclear.

Very little independent research has been done on the technical methods used to protect ORCA data, but the ORCA program is similar to other transit smartcard programs that use RFID tags and onboard card readers to log transactions and to RFID toll cards that enable online toll collection platforms. Research on similar programs has found that information stored on RFID cards is not secure. Because encryption slows down information transfer between cards and scanners and places a computational burden on the system, RFID cards are not encrypted. This means that any device that has NFC capabilities can read transit cards and access whatever data is stored on them. This type of technology has been used to

[116] Sound Transit. "ORCA Privacy Statement." Last modified May 14, 2012. Accessed March 22, 2016. https://www.orcacard.biz/ERG-Seattle-Institution/jsp/static/ORCA%20Privacy%20Statement.pdf, 1.

[117] Sound Transit 2016, 6.

scan EZPass toll fare cards in New York City since 2013[118] as a way of collecting real-time traffic information in congested corridors. In fact, the information stored on ORCA cards is accessible to any smartphone equipped with NFC technology. For example, the app Farebot was designed specifically to access information stored on RFID transit cards and specifically lists ORCA as one of the compatible RFID card systems in their marketing materials. When activated, the app can decode the information stored on a transit pass in the same way that a scanner at a transit station would. While the storage of unencrypted information is a key piece of the system design, it does present a privacy challenge that is outlined in the ORCA privacy policy.

ORCA is governed by a notice and consent policy, which is hosted on Sound Transit's website. It is clear in describing what data is collected and what users consent to by choosing to buy an ORCA pass. It clearly states that the card contains spatial information, and that the system collects personal information in the act of creating an ORCA account and loading value onto the pass. It makes it clear that employers can track where and when particular transit cards are being used and that the unencrypted information stored on the card can be read by non-ORCA devices.[119] It is a good example of a "typical" notice and consent policy, and is clearly meant to read like a legal contract. The privacy statement asserts that no personally identifiable information is stored in conjunction with the card information, with the exception of birthdate and pass type. However, information on pass type and information on the last ten trips taken *is* stored on the card. This is used to

[118] Hill, Kashmir. "E-Z Passes Get Read All Over New York Not Just at Toll Booths." Forbes. Last modified September 2013. Accessed June 6, 2016. http://www.forbes.com/sites/kashmirhill/2013/09/12/e-zpasses-get-read-all-over-new-york-not-just-at-toll-booths/#5e2b61ef3cfc.

[119] Sound Transit 2016, 4.

facilitate ORCA's differentiated payment structure and to allow for transfers, but it also makes it easy to gather information about individuals from their ORCA cards. While the policy itself is rather straightforward, the mechanism of presentation is not. The ORCA privacy notice is not presented when a user purchases a pass nor when they use their pass on a bus or a train. In fact, there is no mention of the *existence* of a privacy policy in any of the materials available to the public at transit stations and pass-selling kiosks. While the policy is easy to find on Sound Transit website, people have to actively look for the privacy policy. This method of (non)presentation is far from the best practices outlined in the literature on notice and consent.

### *Case Study #2: MAC Address Sniffers (Seattle Department of Transportation)*

In the last year, the Seattle Department of Transportation has implemented a WiFi based system of sensors at major intersections in heavily congested corridors in the city. As cars pass through the intersection, the sensor pings all devices with MAC addresses that have WiFi sensing enabled, which includes all smartphones, tablets, and cars with built-in computers. The MAC addresses are tracked through the corridors where sensors are located, and this provides real-time information on the speed, volume, and mode-split of traffic.[120] This data can be useful in planning real-time congestion mitigation strategies, but it also raises privacy concerns both because MAC addresses are highly identifiable information and because there is no current notice delivery protocol in place.

[120] Kroman, David. "Seattle installs new system to track individual drivers." Crosscut. Last modified September 8, 2015. Accessed March 22, 2016. http://crosscut.com/2015/09/seattles-new-technology-tracks-how-we-drive/, 2.

Initial research into the implications of this program were addressed as a part of "Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government". Jan Whittington and a team of multi-disciplinary researchers evaluated the risks and costs associated with eight of the City of Seattle's data collection programs and found that cities share data in three ways: push, pull, and spill. They observe that:

> Cities "push" data when they publish databases through online or other portals. Residents and others "pull" data out of the city with public records requests. And cities "spill" data through accidental exposure, malicious data breach, and the distribution of data by vendors, contractors, and partners.[121]

These categories inform and unify their case studies and help to explain why the MAC address sniffer program is problematic. By working with the Seattle Department of Transportation (SDOT), the research team discovered that data was not only collected without delivery of notice, but that the opt-out option was hosted on the third-party service provider's website (Acyclica.com). Even though this might be the easiest way to manage the opt-out functionality from the city's perspective because it streamlines the integration of the opt-out functionality with Acyclica's software, it makes it even more difficult for people to learn about municipal data collection programs. Because there is no mechanism in place for the delivery of notice for the use of the sensors, the use of the sensors may in fact make residents feel that their privacy has been violated even if the data is managed ethically. That is, the very fact that the sensors are designed to not be noticed and the fact that there is minimal public outreach regarding the program makes people perceive the Acyclica MAC address sensors as a privacy violation.

[121] Whittington et al 2016, 1902.

On a technical level, collection of MAC addresses can present a locational privacy challenge if the system is not carefully designed. In "Unique in the Crowd: The privacy bounds of human mobility," Yves-Aleandre de Montjoye et al. demonstrated that mobility patterns are highly unique and that only four spatio-temporal points are needed to uniquely identify 95% of individuals represented in a city-scale dataset. Using data from cell carriers, the researchers demonstrated that even coarsening the data by aggregating data points into larger geographic areas did little to obscure the uniqueness of the mobility patterns. Specifically, the researchers found that the uniqueness of the patterns decay approximately as the 1/10 power of their resolution, meaning that the level of anonymity provided through aggregation increases at $1/10^{th}$ the rate of the increase in size of geographic area used for aggregation.[122] According to this work, even geographically coarse datasets provide little anonymity. This work is important to the Acyclica case study because it demonstrates that locational data can be used to uniquely identify individuals even when no personal information is collected. In other words, knowing the specific MAC address of a device is not necessary to uniquely identify it. All that is needed to identify a person is a few spatio-temporal points linked via a hashed identifier. Yves-Aleandre de Montjoye's research also demonstrates that typical approaches to designing systems that protect privacy are not sufficient to anonymize datasets. In other words, aggregating data points into larger geographic areas to "cloak" the uniqueness of individual data points cannot be used as a fail-safe way to guarantee individual privacy.

[122] De Montjoye 2013, 2.

De Montjoye's work is also important because it makes it clear that people can be identified even without the collection of personally identifying information (PII). None of the data used in de Montjoye's work is linked to name, gender, or date of birth, but nevertheless, the research team was able to identify people based on their movement patterns. This is important to the Acyclica case study because the Acyclica sensors do not collect any of the types of PII currently protected by law. However, they collect MAC addresses, which uniquely identify each device (e.g., smartphone, laptop, or automotive computer), and those MAC addresses are then encrypted using a proprietary algorithm. The resulting encrypted value is used as a unique identifier that links temporal-spatial points together. The encryption of the MAC address is designed to ensure that data streams are secure and that the data collected cannot be easily combined with other data streams that might lead to privacy violation. However, the fact that all of these encrypted identifiers are still persistent and still connected to individuals makes it possible to uniquely identify people from locational data points.

### *Case Study #3: Automatic Licenses Plate Recognition (Seattle Police Department)*

The Seattle Police Department, like other urban police forces across the nation has an Automated License Plate Reader (ALPR) data collection program that uses cameras linked to computers to scan license plates and check them against a database of offenders, and log the location, date, and time of the sighting. The department uses the data to find stolen cars, write overtime parking tickets, and assist in ongoing criminal investigations. Although vehicle registration data is deemed private by Washington state law, the license

plate numbers themselves and the geographic coordinates are not.[123] This is a locational privacy risk because anyone can file a public records request and obtain the collected data at the database level. That is, the Seattle Police Department is required by law to disclose ALPR data when it is requested and cannot obscure license plate information in any way.

Because the license plate numbers are not encrypted, it is possible to track the movements of a single vehicle through the city and to link the license plate number to other records in other databases. Because license plate information is used in background checks, credit checks, and a myriad of other non-governmental databases, it is possible that ALPR data could be cross-referenced with other databases to build detailed and specific profiles on individuals based on their patterns of movement throughout a city. This argument, known as the "linkage argument", has been a recurring thread in the work of Helen Nissenbaum,[124] Adam Moore,[125] David Kroman,[126] and countless others. As de Montjove demonstrated in "Unique in the Crowd," four spatio-temporal data points are enough to uniquely identify an individual in a city-scale dataset; when that level of granular data is combined with the level of information that is commonly tied to a vehicle registration, it becomes evident that ALPR databases are a large source of informational privacy concern.

---

[123] Seattle Police Department. "16.170 - Automatic License Plate Readers." Seattle Police Department Manual. Last modified June 2012. Accessed March 22, 2016. http://www.seattle.gov/police-manual/title-16---patrol-operations/16170---automatic-license-plate-readers.
[124] Nissenbaum 2010, 146.
[125] Moore, Adam, ed. Privacy, Security, and Accountability: Ethics, Law, and Policy. N.p.: Rowman & Littlefield International, 2015, 5.
[126] Kroman, David. "Seattle installs new system to track individual drivers." Crosscut. Last modified September 8, 2015. Accessed March 22, 2016. http://crosscut.com/2015/09/seattles-new-technology-tracks-how-we-drive/.

Another challenge that ALPR systems present for locational privacy is that the technology has no clear way to deliver notice and provide an opportunity for drivers to consent to the program[127]. Short of placing notices on every police car equipped with ALPR technology, there is no way to let people know that their license plates are being recorded in real time. Even if the police department were to place notices on police cars, the signs would likely be too small to read at a typical driving distance. Because the ALPR technology scans (or attempts to scan) every license plate, it is not possible for people to opt-out of the data collection. The design of the technology makes this impossible.

Between the lack of notice and consent, the lack of encryption on the data collected, and the legal environment of Washington State, it is clear that the ALPR technology used by the Seattle Police Department presents locational privacy concerns that deserve further scrutiny. This thesis endeavors to gather more information on the system design of the ALPR technology and on specific policies that effect the collection and use of ALPR data.

## Research Gaps

Although informational privacy has many proponents, work still remains to be done on finding ways to protect informational privacy in the smart city. The discussion above about the challenges of protecting locational privacy in a wide range of data collection platforms demonstrates that privacy considerations need to be an element of every data collection program and that the ways in which privacy is protected in practice

[127] Dryer, Randy L., and Stroud, S. Shane. "Automatic License Plate Readers: An Effective Law Enforcement Tool or Big Brother's Latest Instrument of Mass Surveillance? Some Suggestions for Legislative Action." *Jurimetrics Journal of Law, Science and Technology* 55, no. 2 (2015): 225.

can vary widely based on data types, system design, and legal mandates. While the general shape of privacy protection can be inferred from existing literature on the case studies considered above, the devil is in the details. There is no exhaustive body of research that delves into the minutiae of system design and policy management for any of the programs discussed above. There is more work to be done in understanding the challenges that municipalities face when tasked with collecting, managing, and using data in departments across the city. That is, there is a deficit in research on privacy protection at the local level. This is especially true for smart city data collection programs; of all of the sources used in the literature review, only four addressed the ways in which cities are addressing privacy protection in smart city sensor networks.[128] This thesis contributes to the field by highlighting the role of locational data in compromising privacy, and conducting case studies of smart city sensor programs in the City of Seattle.

[128] Whittington et al 2016, Schlaub et al 2015, Shvartzshnaider et al 2016, and Finch and Tene 2015,

# 3. Methods

This research involves exploratory case studies of municipal agencies in Seattle that use distributed sensor networks to collect data about behavioral patterns of city residents. Data was gathered through publicly available documents, analysis of existing case studies, and interviews with primary decision-makers in charge of smart city sensor programs. Through this research, I hope to answer the research questions posed at the beginning of this paper:

1. What are the perils and promises of location-based smart city data collection programs?
2. What are cities with smart city data collection programs currently doing to protect locational privacy?
3. How can current/existing practices for protecting locational privacy be improved?

## Research Design

This research study takes the form of a multi-part exploratory case study. A multi-part exploratory case study was selected in order to evaluate different locational privacy protection strategies that have been implemented in Seattle's municipal agencies, which are actively embracing smart city technologies. The multi-part exploratory case study approach is appropriate because each agency has prioritized different aspects in the implementation of smart city data collection programs and must be evaluated as a holistic unit before it can be compared to other agencies.[129] Particular attention will be paid to the context of use, which is particularly important in exploring the application of information technologies.[130] The replication logic of a multi-part case study exploratory model will

[129] Tellis, Winston. "http://www.nova.edu/ssss/QR/QR3-3/tellis2.html." The Qualitative Report. Last modified September 1997. Accessed March 22, 2016. http://www.nova.edu/ssss/QR/QR3-3/tellis2.html, 4.
[130] Tellis 1997, 2.

provide multiple sources of information with which to evaluate the strengths, weaknesses and opportunities of current privacy practices in smart city data collection programs.

The unit of analysis[131] used in the case study is the individual smart city sensor program. While a full implementation of a smart city network would require data streams from multiple sources, many city agencies are implementing smart city data collection programs one at a time, in hopes of reaping the benefits of city-level ICT while minimizing upfront costs for new infrastructure networks.[132] Also, each sensor program is structured differently and will have different approaches to protecting locational privacy. By examining each data sensor program individually, we can see both fine-grain technical privacy solutions and high-level policy decisions that are intended to protect privacy.

The comparison of three smart city data collection programs will illuminate issues related to privacy and data collection and will show a range of responses to the problem. The issues raised by these case studies include concerns about the usefulness of data, information security, regulatory compliance, and individual privacy. Interviews with key decision makers in charge of implementing each smart city program delve into perceived concerns about new data collection methods, exploring the lessons learned from implementing privacy-sensitive technologies and handling both open data imperatives and public records requests for large and costly datasets. The questions asked during these interview are designed to provide an open-ended framework that will elicit a deeper understanding of privacy concerns and protection efforts that are specific to each program

[131] Yin, Robert K. *Case Study Research : Design and Methods*. 3rd ed. Applied Social Research Methods Series ; v. 5. Thousand Oaks, Calif.: Sage Publications, 2003, 46.
[132] Finch and Tene 2014, 2.

while still drawing out the commonalities between the different cities and different programs.

## Interviews: Experience and Opinion

Interviews with decision-makers in smart city programs and information and communication technology professionals will supplement the case study review. These interviews will offer a wealth of information garnered through decades of experience implementing and managing technology systems in the urban landscape. The purpose of these interviews is to better understand how key decision makers understand the risks and benefits of smart city sensor programs.

The interviews will be conducted with project managers and department heads in charge of each urban data collection program outlined above. The interview protocol will involve asking a series of open-ended questions designed to gather information about experiences with implementing smart city sensor programs and to develop a working understanding of how each program works to protect locational privacy of data subjects. Although the questions are open-ended, they will be tailored to the specifics of each program in order to have productive and meaningful conversations. The questions will focus on locational privacy, notice and consent, data management, data security, and privacy regulations.

While not intended to lead to a statistical analysis, the interview data will provide information on program manager's perceptions on the importance of privacy protection in the complex environment created by smart city programs. In conjunction with the background research discussed in Chapter Two and reviews of publicly available

documents related to each sensor program, the interviews will provide information on areas of concern in smart city sensor programs as well as highlight areas of agreement and divergence amongst the program managers. These multiple sources of data will serve as a method of triangulation by which to build new knowledge.[133]

## Interview protocol

### Introduction:

"This is research for my Master's thesis. Your names will be confidential, but I will likely use the agency name when I talk about the technology. If at any time you want to end the interview, or not answer a question, that is quite alright. You can also tell me if there is something you want to be off the record."

"I am interested in the kinds of data you are collecting. I am looking at 'notice and consent,' the practice of letting people know when their data is collected, and how that breaks down in practice when designing large sensor systems."

"What is your role in working with the sensor network, and what responsibilities does it entail?"

"I am interested in the kinds of data the program collects for the city of Seattle. Will you please walk me through, step by step, how this data is collected, what specific attributes are in that data at each step as it moves between the sensors, the third party contractor, and the city? "

"Part of my research includes a side-by-side comparison of data attributes that are being collected by different sensor networks. Would you be willing to provide me with a list of data attributes that are collected by the sensor system?"

### Privacy

"What works well about this program with respect to resident privacy?"
"What would you like to see improved?"
"To what extent was the current system shaped by local and national privacy laws and regulation?"

### Notice and consent

"Are you familiar with notice and consent?"
"Is the delivery of notice to data subjects part of your operational model?"
"(If so) how is it delivered? OR (If not) why not?"
"Are people asked to consent to data collection? If they are, can you describe the method in which they are asked for consent?"

[133] Tellis 1997, 16.

"If notice and consent are a part of the operational model, how are they put into practice? If they are not, why not?"

**Data management**

"How is data from the sensors managed?"
"Is data from the sensors released to the public through an open data portal?"
"Are you aware of any other purposes for which the third party contractor uses the data that it collects?"

**Data security**

"What measures have been take to prevent hacking of smart sensor data streams?"
"What auditing policies are in place to ensure regulatory compliance?"

**Closing questions**

"Is there anything I haven't asked about that you think would be relevant?"
"Is there anything you'd like to add?"
"Is there anyone else you think I should talk to?"
"Do you have any questions for me?"

### Interpretation of Results

The results of the case studies will be presented through a two-pronged Strengths, Weaknesses, Opportunities, and Threats analysis (SWOT). The first part of the analysis will present the best practices resulting from the literature review and demonstrate how the concepts from the literature could be used to inform policy and design decisions. The second part of the analysis will reflect the actual policy and design decisions uncovered through the interview process. The differences between these analyses will be discussed in a gaps table, which will highlight the successes and shortcomings of current practices for each case study. Illustrative maps will be provided to give a sense of the scope of data collection programs.

### Limitations and Constraints

As with any qualitative research study, this multi-part case study design has important limitations including interviewer bias; the difficulty of generalizing conclusions from contextualized analysis of agency-level programs; and the limitations of analyzing qualitative data. To reduce interviewer bias, the questions were designed to be open-ended

but specific. They were designed to reduce the potential that interviewees would be likely to shape answers to suit the thesis outlined in this paper. In addition, opinions expressed by previous interviewees were kept confidential, in the attempt to reduce outside influence on each individual interviewee. While it is not possible to create statistical generalizations from case studies[134], the goal of this work is to provide illustrative examples of the ways in which different sensor systems can both create and solve the problem of locational privacy protection.

The integration of information from documentation, interviews, and scholarly research will serve to create a holistic approach to understanding the ways in which locational privacy can be protected in the real world. This triangulation of gathering data through several sources[135] will help to generate a more complete picture of locational privacy protection in smart city programs than currently exists.

[134] Yin 2003, 38.
[135] Tellis 1997, 12.

# 4. Results

The results of the case study methodology reveals specific policy and system design considerations for municipalities using sensor networks to collect data. The case study interviews were connected to the theoretical literature through the use of pre and post interview SWOT analyses, which display perceived strengths, weaknesses, opportunities, and threats. The pre-interview SWOT analysis reflects the best practices derived from the literature review, while the post-interview SWOT analysis reflects the actual concerns of information managers working directly in the programs. Results from each SWOT analysis are compared in a gap table. The results section also includes a table reflecting the data attributes collected by each sensor network and the intended purpose of the data. Hopefully, the two-part SWOT analysis reveals the need to consider both design and policy in the implementation of smart city data collection programs. The summary of how the case studies answer the research questions is presented below:

1. What are the perils and promises of location-based smart city data collection programs?

Location based data collection programs deliver high-quality data that can be used to expedite city planning processes. In many cases, city services that we have come to see as essential depend on the use of aggregated location-based data. There is much to be gained from this technology, and it is likely to become more heavily used in the future. That being said, location data is perilous if it is not carefully handled. Spatio-temporal data can be used to develop detailed profiles on individuals and can be linked

to other datasets, leading to new, inferred information that might be of a sensitive nature.

2. What are cities with smart city data collection programs currently doing to protect locational privacy?

For the most part, the case study data collection programs rely upon a combination of privacy by design and back-end anonymization and encryption techniques to try to protect privacy. None of the programs had a notice regime, and only one program (Acyclica) provided an opportunity to opt-out of data collection. All of the agencies said that they were not interested in attempting to provide notice of data collection at the time of collection, and interviewees consistently stated that the information about the programs available online was sufficient to inform people about the data collection programs. With the exception of the Seattle Police Department, all of the interviewed agencies used third party contractors to manage data flows and provide data aggregation services.

3. How can current/existing practices for protecting locational privacy be improved?

This question will be addressed more fully in the next chapter, but one recurring theme from the case study research was that the Washington State Public Records Act is so broad that almost any data collected by a government agency can be requested. This was a major concern to privacy advocates in city agencies and was consistently rated as the primary barrier to improving privacy practices in the agency.

Clearly, location-based sensor networks provide invaluable data to city agencies. However, these programs also pose privacy risks. These concerns are evident in the two-prong theoretical and experiential analysis of each program's strengths, weaknesses, opportunities, and threats (SWOT) presented below. Table 1 Data Collected and Stated Purposes for Programs shows the types of data collected by each program, Table 21 shows the results of the literature-based SWOT analysis, Table 32 presents the results of the interview-based SWOT analysis, and Table 43 analyzes the gaps between the two SWOT analyses. Maps representing data collection locations for each program are included, and further discussion of the results will follow.

*Table 1 Data Collected and Stated Purposes for Programs*

| Technology | Data Collected | Information Directly Inferred |
|---|---|---|
| **Acyclica Sensor Network** | MAC addresses | Speed through arterial corridors |
| **Automated License Plate Readers** | Vehicle license plate numbers | Vehicle's location and association with crimes or wanted persons |
| **ORCA Smartcards** | Pass type, pass price, employer id, origin bus stop, destination bus stop, e-purse value, dates of last ten trips | Validity of pass for payment transaction |

*Table 2 Pre-Interview SWOT Analysis*

| Technology (Research) | Issue | Strength | Weakness | Opportunity | Threat |
|---|---|---|---|---|---|
| **Acyclica Sensor Network** | **Data Collection** | Data should be aggregated and all identifiers for trips should be discarded. | System relies on tracking MAC addresses from individual phones; nodal aggregation may not be possible. | Further enhancements of privacy by design. | Trip data can be used to identify individuals, predict future behaviors, and determine future locations. |
| | **Data Retention and Management** | Data should be retained only in aggregated form. All disaggregated data should be destroyed. | Persistent identifiers are assigned to individual trips to calculate traffic speeds. Identifiers may be used to track individuals. | Data minimization, limiting data retention to 90 days for both Acyclica and the City of Seattle | Risk of hacking, encryption failures, or other unintentional data release. Use of disaggregated data for commercial purposes. |
| | **Public Outreach** | SDOT should actively seek to inform the public about data collection, management, and retention practices. Transparency is key. | There is no current mechanism for notification at the time of data collection. | Use privacy beacons to "piggyback" on WiFI sensors. Phones could receive notification of data collection even as they provided information to the sensor network. | Public backlash when the data collection program becomes public knowledge through misuse of collected data. |
| | **Privacy Regulation** | Only the aggregated data accessible by SDOT is subject to PRA. | PRA is not clear about whether or not data collected by a third party on behalf of a city is subject to disclosure. | PRA can be amended to specifically allow aggregated location data release but protect disaggregated data. | Wide scale release of locational data through the use of the Public Records Act for commercial purposes, or uses that are not in the public interest. |

| Technology (Research) | Issue | Strength | Weakness | Opportunity | Threat |
|---|---|---|---|---|---|
| **Automated License Plate Readers** | **Data Retention and Management** | Data should be encrypted, accessible only to police officers serving a warrant, and retention should be limited to positive matches rather than all scans. | Encrypting data and removing non-matches may require extra processing power or data cleaning. | Provide high quality law-enforcement while still protecting privacy. Remove threat of large-scale behavioral analysis based on this data. | Analysis of individual and collective movement patterns over time through wholescale release of unencrypted data. |
| | **Privacy Regulation** | ALPR data is subject to the Public Records Act. | The Public Records Act has no exception for locational data. | Provides strong rationale for reworking the PRA to address the challenges of modern technology. | Wide scale release of locational data through the use of the Public Records Act. Lack of control over downstream uses of the data. |
| | **Open Data** | Aggregation may prove sufficient to protect locational privacy while allowing for transparency. | Disaggregated ALPR data should not be released through an open data portal; doing so would create locational privacy risks for individuals. | Study aggregation techniques suited to spatio-temporal datasets; create method of release that are privacy-aware. | Wide scale release of disaggregated locational data. Analysis of individual and collective movement patterns over time through wholescale release of unencrypted data. |
| | **Public Outreach** | Notice should be given both online and at the time of data collection. | Notice cannot currently be delivered at the time of data collection due to technological constraints. | Stakeholder outreach, improvements in privacy by design, or reworking of PRA. | Massive public backlash when people realize how SPD is tracking their movements. |

| Technology (Research) | Issue | Strength | Weakness | Opportunity | Threat |
|---|---|---|---|---|---|
| **ORCA Smartcards** | **Data Retention and Management** | Data on cards should be encrypted and only accessible via an authentic card reader. Data should be managed in-house. | Encryption and in-house data management impose costs on the agency. | In-house data management, research into streamlining encryption techniques, and advances in privacy by design. | Large-scale data release, targeting of vulnerable populations, and threat of hacking. |
| | **Anonymization and Aggregation** | Data on cards should be encrypted. Research data should be anonymized and aggregated. | Loss of usefulness of research data, costs in time due to encryption lag in fare payment transactions. | In-house data management, data minimization, research into encryption techniques. | Large-scale data release, targeting of vulnerable populations, and threat of hacking. |
| | **Privacy Regulation** | ORCA data is controlled by ORCA privacy policy, and is only given out to employers and researchers. | No codified data access protocol. | Development of clear data access procedures, perhaps some form of data access for the public. | Unclear if data is subject to PRA. |
| | **Ensuring Privacy** | ORCA has developed a privacy policy and posted it on its website. | Notice is not be delivered at the time of data collection. | Public outreach, posting of notices, or privacy by design (encryption of RFID chip). | Massive public backlash when people realize how Sound Transit is tracking their movements. |

*Table 3 Post-Interview SWOT Analysis*

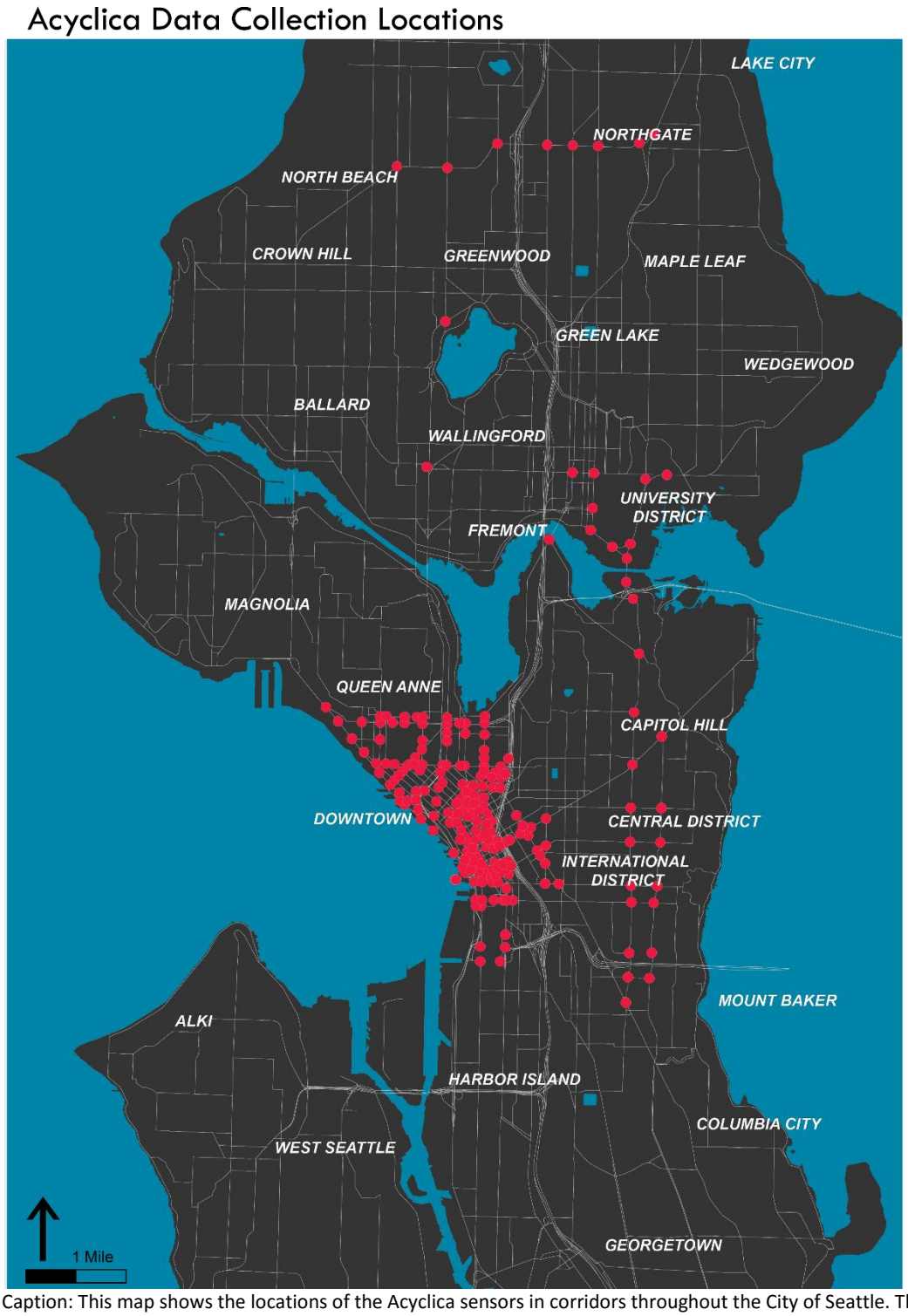
| Technology (Interview) | Issue | Strength | Weakness | Opportunity | Threat |
|---|---|---|---|---|---|
| **Acyclica Sensor Network** | **Data Collection** | Acyclica sensors collect detailed, granular data in ways not possible with other sensor networks. | Data collected at the person level. No way to measure traffic flows through a corridor without also recording individual phone signals. | Continue to improve privacy by design features like encryption, data minimization, and immediate aggregation. | Hacking or spills from databases. Data requests under PRA. |
| | **Data Retention and Management** | Data is maintained in an encrypted form on third party servers. Only aggregated through-put through arterial corridors is recorded by the city. | Persistent identifier still required to compute traffic speeds for corridor segments. Data is used for real-time congestion mitigation and temporal analysis. | Retain as little data as possible given legal and computational requirements. | Hacking or spills from databases. Data requests under PRA. |
| | **Public Outreach** | Limited news coverage already complete. Agency employees willing to answer questions. | No information about Acyclica sensor network is available on SDOT's website. No privacy policy or notice form available. | Creation of a privacy policy or data collection notice. | Public backlash against data collection. |
| | **Privacy Regulation** | Data is subject to PRA, but only the aggregated summaries delivered to SDOT by Acyclica. | Court-ordered data release could target de-aggregated database. | Continue to improve privacy by design features like encryption, data minimization, and immediate aggregation. | Hacking or spills from databases. Data requests under PRA. |

| Technology (Interview) | Issue | Strength | Weakness | Opportunity | Threat |
|---|---|---|---|---|---|
| **Automated License Plate Readers** | **Data Retention and Management** | ALPR data is subject to PRA in its entirety. | Vehicle license plate numbers can be used to link datasets and to trace individuals through the city. It can also be used to predict time-based behavior patterns. | Encryption of data would make it more difficult to mine. | Largescale linkage of datasets that leads to revelation of personal information and invasion of privacy. |
| | **Privacy Regulation** | PRA is designed to ensure government transparency. | PRA makes no exception for the linkage argument; argues that there is no right to privacy in public spaces. | Revision of PRA to update for new technological capabilities. | Largescale linkage of datasets that leads to revelation of personal information and invasion of privacy. |
| | **Open Data** | Open data has the potential to reduce PRA requests and to provide crucial information to citizens about government operations. | In practice, open data generates more PRA requests because data intended for the public is often aggregated or limited. Open data piques interest; PRA is an easy way to obtain complete datasets. | Use Open Data Portal to provide information, reduce number of PRA requests, and inform the public about data collection. May help quell transparency concerns related to changing PRA. | Cost and work of cleaning data for Open Data portal, generation of new PRA requests for disaggregated or more complete data. |
| | **Public Outreach** | SPD hosts a legal statement about ALPR on website and talked to reporters in the past about the program. | Public outreach campaign not a priority, no privacy policy, no ability to opt out, notice not delivered at time of data collection. | Creation of a privacy policy or data collection notice. | Public backlash against data collection. |

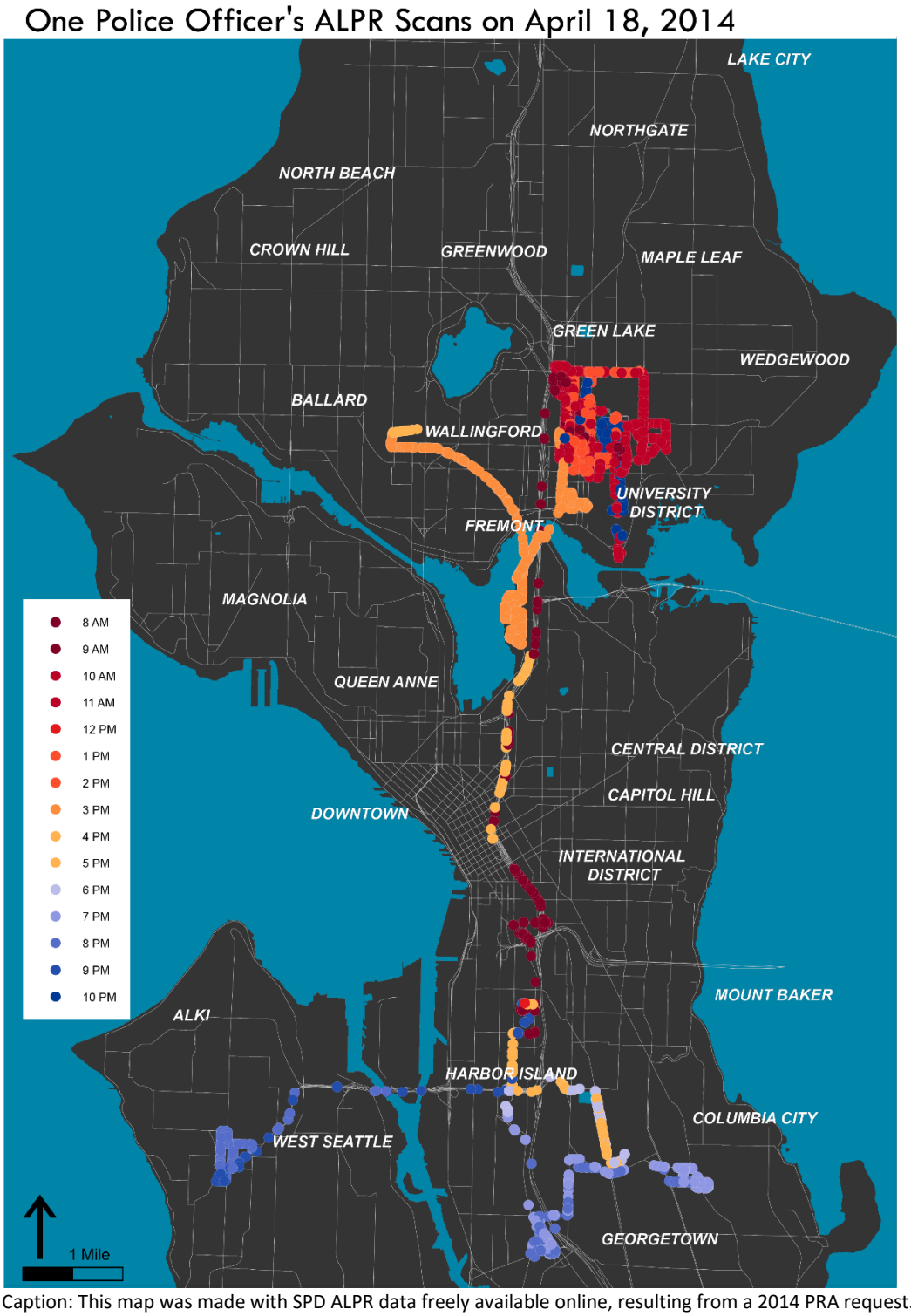
| Technology (Interview) | Issue | Strength | Weakness | Opportunity | Threat |
|---|---|---|---|---|---|
| **ORCA Smartcards** | **Data Retention and Management** | Data is not subject to PRA, Sound Transit requires IRB approval and data privacy plan before releasing data. | No formal data release plan; decisions are made on a case-by-case basis. | Creation of a process for obtaining data access. Creation of a secure database for researchers. | Data spills, creation of detailed profiles of individual customers. |
| | **Anonymization and Aggregation** | Data attributes can be stripped from records. Also, data is only accurate to bus stop level, which is geographically opaque. | The full set of data attributes collected could be used to trace individuals. Data attributes include pass type, employer, pass price, and stops. | ORCA Card encryption, de-identification, and aggregation. | Data spills, creation of detailed profiles of individual customers. |
| | **Privacy Regulation** | Data is handled by Vix, so Sound Transit is not liable for data breach or database maintenance. | Vix has not been a reliable contractor and has repeatedly cost Sound Transit time and money through poor system design and data management. | Selection of new contractor with ORCA2, with a focus on ensuring that the new contractor is complying with privacy policy and is not using data for commercial purposes. | Data spills, creation of detailed profiles of individual customers. |
| | **Ensuring Privacy** | After initial feedback on data security, Sound Transit is now very aware of privacy concerns and determined to implement solutions. | Data is not easily accessible. Only one research team to date has access to full dataset. | Create a streamlined process for research verification and data release. Look to Census data structure for inspiration. | Unnecessary data blocking slowing down transparency efforts and innovation from researchers and civic hackers. |

*Table 4 Gap Analysis Revealing Differences in SWOT Analyses*

| Issue | Research-Based SWOT | Interview-Based SWOT |
|---|---|---|
| **Data Collection** | ○ Data collection should be minimized as much as possible, should be linked to a specific service delivery.<br>○ Encryption and aggregation should be used whenever possible. | ○ Data collection linked to specific service delivery.<br>○ Encryption not used if transaction speed is affected. |
| **Data Retention and Management** | ○ Data should be retained only in aggregated and anonymized formats.<br>○ Data retention should be limited to a specific length of time. | ○ Data is stored in unencrypted databases and held by third party contractors.<br>○ Data is retained indefinitely. |
| **Public Outreach** | ○ Notice should be delivered at the time of data collection.<br>○ Privacy self-management options should be provided through system design. | ○ Public outreach not major concern.<br>○ Interviewees not concerned with informing public about data collection practices. |
| **Privacy Regulation** | ○ Locational privacy should be protected via law and policy. | ○ All data studied (with the exception of ORCA fare transaction data) was subject to all-inclusive public records requests. |
| **Open Data** | ○ Data should be anonymized and aggregated before being made public online. Open data is a good way to create transparency without jeopardizing privacy. | ○ Publication of aggregated or redacted data through open data portal leads to an increase of PRA requests for complete datasets. |
| **Anonymization and Aggregation** | ○ Anonymization and aggregation are necessary but not sufficient to protect privacy. | ○ Data aggregation is done after data collection. Data anonymization only removes PII, not all identifiers. |
| **Ensuring Privacy** | ○ Privacy concerns should be addressed by agencies, not individuals. | ○ Privacy self-management is still predominant technique. |




Caption: This map shows the locations of the Acyclica sensors in corridors throughout the City of Seattle. There are currently over 200 sensors in major arterial corridors in the city, and SDOT is interested in expanding the program. Program managers believe that this is one of the most successful traffic mitigation technologies available.




Caption: This map was made with SPD ALPR data freely available online, resulting from a 2014 PRA request. The original dataset had over 1 million rows, so I subset the data to show only the vehicle license plate scans from one police car on April 18, 2014. This map alone should not be taken as a quantitative analysis; instead it is a demonstration of the power and availability of this data.

ORCA Data Collection Locations


Caption: This map shows the locations at which ORCA card data is scanned throughout the city. While it is evident that data is collected throughout the city, it should be noted that the fact that data is collected at bus stops provides a level of geographic obscurity, making it difficult to determine precisely where someone is coming from or going to. Thus, nodal aggregation is built into the data collection methodology.

## Inferences from SWOT Analysis

The two-prong SWOT analysis and the subsequent gap analysis reveal that many of the concerns that are considered in the theoretical literature have yet to be implemented by municipal agencies in the City of Seattle and at Sound Transit. This was especially noticeable for notice and consent principles; while all three agencies interviewed were already familiar with the principles of notice and consent, none provided notice at the time of data collection or actively sought to encourage people to engage with available opt-out options. Each agency clearly stated that its first goal was to provide high-quality, cost-effective service to the people of Seattle and that the use of these data sources was a critical element in their ability to achieve that goal. They made it clear that in most cases, the perceived benefits of using location-based sensor networks was higher than the perceived costs. While all three agencies were keenly aware of the need to protect individual privacy, none of them believed that it was in the best interest of the public to change policies and practices to better protect privacy. In cases where privacy was a concern, program managers relied heavily on either privacy by design principles or an internal IRB process to control data collection and release. While these methods may prove to be effective, there is still room for improvement in privacy protection in each of the case studies.

While the gap analysis provides a summary overview of the differences that emerged through the interview process, the full breadth of the issues unique to each technology must be considered individually on a case-by-case basis.

## Limitations and Constraints

The limitations of this research process include interviewees' differing levels of familiarity with privacy literature and regulatory frameworks. This led to differing levels of focus on the privacy implications of each technology versus the technological specifications of the platform. While both components are necessary parts of a system design, no one interviewee was intimately familiar with both aspects. Additionally, as current employees of city agencies, the interviewees may have felt obligated to defend current city data collection practices. Finally, it would have been useful to interview more people at each agency and to include a broader range of cities in the research project.

### *Acyclica Sensor Network*

The two-pronged SWOT analysis for the Acyclica sensor network revealed several divergences between the theoretical literature and the actual practices of the agency. It also revealed that the program relies primarily on privacy by design techniques to try to protect individual privacy. In the theoretical SWOT analysis, the main concerns were locational privacy, data management, privacy regulation, and notice and consent. In the interview-based SWOT analysis, the main issues were data collection, retention and management, public outreach, and privacy regulation. The interviewees made it clear that they believed that the potential privacy implications of the sensor network were minimal and that the accurate traffic congestion data was revolutionary in SDOT's work to mitigate congestion in arterial corridors. The following sections present findings and discuss implications.

## Data Collection

The Acyclica sensor network depends on MAC address sniffers installed on traffic signals in major arterial corridors. The sniffers "look" for smartphones that have their WiFi capability turned on. When the sensor finds a phone with an activated WiFi signal, it

records the MAC address of the phone and assigns it both a date-time stamp and a salted identifier and discards the actual MAC address. The salting method is a proprietary encryption method that has been verified through a Seattle Department of IT (DoIT) process to ensure that the data protection method meets city specifications. The salted identifier is persistent, meaning that when another sensor in the network senses the same phone at another location, it will be assigned the same salted identifier. By putting the records for a certain identifier in chronological order, Acyclica can assess the speed of traffic in corridor segments at any time of day or night. While data is gathered at the person-level, information on individual trips is not available to SDOT engineers. Acyclica aggregates all data at five-minute intervals and displays only the aggregated summation of traffic speeds by corridor segments in its city-facing software platform. While this is done to reduce dataset complexity, it has the side effect of ensuring that individual trips cannot be tracked through SDOT's version of the data.

### Data Retention and Management

Data is carefully cultivated to ensure that the traffic engineers at SDOT are able to complete annual and quarterly studies of average traffic speeds in arterial corridors. This data is retained in aggregate form. That is, only the segments summaries are recorded. All the disaggregate data, including MAC addresses and salted identifiers for individual trips are discarded and aggregated before the data is delivered to SDOT or recorded in the archive database. SDOT traffic engineers access Acyclica data through an online web portal. It provides the ability to query by spatial location, time of day, or level of congestion. It relies primarily on a map-based interface to allow engineers to select routes

of interest, but it also has the ability to generate static reports that provide detailed information about the amount of time it took to travel a particular segment. Variations in speed allow Acyclica sensors to provide accurate "guesses" about what mode individuals are traveling by. Thus, the Acyclica data is capable of providing quantitative data for multimodal analysis, which has never been possible before on a large scale. Data points associated with different modes are retained and flagged for multimodal analyses.

The long-term retention of corridor segment speeds is useful for analyzing the effectiveness of congestion mitigation treatments, assessing the relative state of traffic congestion in the city, and in performing econometric modeling of the effects of traffic congestion on Seattle's economy. Because data is stored in aggregated form and is not data that could easily be disaggregated, SDOT believes that there is no privacy risk to retaining historical traffic speed data generated by the Acyclica program.

### Public Outreach

Although SDOT has been willing to talk to reporters[136] and has participated in past research on municipal privacy practices[137], the interviewees stated that they did not see a reason for conducting wide-spread public outreach or for providing notice at the time of data collection. The interviewees stated that sufficient public outreach about the technology had already been conducted because of the past news articles and that the nature of the technology makes it nearly impossible to deliver notice. Delivering notice and requesting consent from every person to travel through the busiest corridors in Seattle is a Herculean task. If it were required on a daily basis, it would create an unnecessary

[136] See Kroman 2015.
[137] Whittington et al 2016, 1899.

annoyance for travelers and a massive computational burden for the city. Thus, SDOT has no intention of attempting to use a traditional notice and consent framework to alert people to data collection practices.

However, the Acyclica sensor network does comply with Seattle's new privacy policy because it provides privacy-aware travelers the chance to opt-out of data collection. People who prefer to not be tracked by the sensors can enter their MAC addresses into a tool hosted on Acyclica's website. Acyclica then uses the list "opt-outs" to exclude those MAC addresses from data collection any time that they are sensed by a sensor. However, finding the opt-out option requires careful searching through Acyclica's website; it is not mentioned anywhere on SDOT's website. The lack of information about the data collection program and the fact that the opt-out option is not made readily accessible may simply be oversight, but it is a needless barrier to privacy self-management. While it may be impossible for SDOT to provide notice to travelers when their MAC addresses are recorded by Acyclica sensors, it would be easy for the agency to develop materials informing the public about data collection efforts and providing clear directions on how to opt-out of data collection.

### Privacy Regulation

The Acyclica sensor network data is subject to PRA; any data collected by the agency in connection to the Acyclica program can be requested through PRA. The interviewees were not overly concerned with public records requests because only aggregated information is accessible to SDOT employees. That is, all MAC addresses are discarded and individual trips are combined based on location and trip direction before Acyclica sends data to SDOT. Several PRA requests for Acyclica data had already been received.

Depending on the breadth of the request, city employees gathered datasets, emails, reports, and analyses and emailed them to the person who had placed a request. To prepare for this eventuality, all data and corresponding documentation is stored in project folders, which can be easily shared with people who have placed PRA requests.

In many respects, PRA is not a concern for Acyclica program managers because the system is designed in such a way that the city never sees disaggregated data. This is a clear application of successful privacy by design; citizens are able to access aggregated data through public records requests, but the data that is distributed in accordance with state law is not compromising to individual privacy. Indeed, the aggregation by five minute intervals into segment-level data is useful to traffic engineers and to citizens interested in using the data to generate publicly-available traffic information. In this way, privacy by design principles have been used to effectively protect locational privacy while also reducing the complexity of the data released to interested members of the public.

## New Innovations in Privacy by Design

Of the three case studies reviewed in this research project, the Acyclica sensor network may be the best positioned to reap the benefits of new research in privacy technologies. For example, Konings et al.'s research on Prifi beacons, which broadcast privacy implications of sensor networks using the same MAC address signaling as a typical device-to-router communication.[138] While this research only demonstrated the possibility of providing notice of data collection and did not address any sort of opt-in/opt-out functionality, it is still a step in the right direction. Perhaps, if this technology were used in conjunction with

[138] Konings et al 2013, 85.

the Acyclica platform, people could be directed to a website where they could choose to opt out of future data collection. While this research has yet to be tested in the real world, it clearly demonstrates that privacy by design and notice and consent can be used in conjunction to help people better understand the workings of the smart city.

### *Automated License Plate Readers*

Automated License Plate Readers (ALPR) are by far the oldest of the technologies studied in this research effort. Indeed, ALPR technology has been available since the 1970s and has been in widespread use since the 1980s.[139] Even though the technology is well-established, privacy measures designed to ensure that data generated by the system are used responsibly have not kept pace. Because ALPR works by taking a picture of a license plate, scanning the numbers in the picture into a machine-readable database, and looking for matching license plates in a database of wanted vehicles, encryption mechanisms are impractical and are not a part of the system design. According to industry experts, encryption would place large computational burdens on the system and would require that all databases used for record matching use the same encryption algorithm. Creation of a persistent identifier for vehicle records would require high levels of interagency cooperation as well as a high level of technical competency and is deemed impractical by the industry. This means that the license plate readers record and store the license plate number of every vehicle that they scan.[140]

---

[139] Dryer and Stroud 2015, 2.

[140] While there have been recent innovations in privacy aware cameras, the current versions of the technology require that each individual (or in the case of ALPR cameras, each individual vehicle) have a privacy beacon that broadcasts privacy preferences to nearby sensors. This technology has not been tested outside the lab and would require such a large shift in multi-sector infrastructure that it is difficult to imagine that it would be adopted in the near future. Thus, privacy by design solutions for ALPR systems are not viable for near-term privacy protection.

When the report logs of ALPR sensors are aggregated into a single database, the technology generates an incredibly rich data source that can track the movement of almost any vehicle through the city in a given day. ALPR data is used to find stolen vehicles, track wanted persons, or aid in a variety of other police activities. While the technology is not imperfect, it is without a doubt one of the best tools that the police department has to conduct reconnaissance missions. Although the department is keenly aware of the security risks of the technology, the fact remains that it is incredibly useful and that there is no other available alternative that provides the same service while still protecting individual privacy. Thus, the police department intends to continue to use ALPR sensors for the foreseeable future. The following sections present findings and discuss implications for this case study.

## Privacy Regulation and Data Release

The PRA has deep and unavoidable consequences for ALPR programs in Washington State. The PRA states that all written records of government agencies are subject to public records requests unless there is a specific exemption for a particular type of record. The PRA makes no exemption for information related to privately owned vehicles, so the Seattle Police Department is required by law to release the entirety of the records that it collects. SPD has received several requests for all ALPR data and has complied with these requests by delivering databases containing ALPR data for a three month period. The data has been used for at least three doctoral level research projects[141] and has perhaps been

[141]Newell, Bryce Clayton. "LOCAL LAW ENFORCEMENT JUMPS ON THE BIG DATA BANDWAGON: AUTOMATED LICENSE PLATE RECOGNITION SYSTEMS, INFORMATION PRIVACY, AND ACCESS TO GOVERNMENT INFORMATION." Maine Law Review 66 (2014): 397-591.

incorporated into privately-held data warehouses. At the time of writing, a 2014 dataset of four months of license plate scans was freely available online, the result of a public records requests. The dataset consists of two tables of a million records each, and records information like license plate, time, location, police unit, and whether or not there was a record match with any vehicle of interest. The fact that this data is easily accessible online drives home the point that PRA requests can have important privacy implications for millions of people.

While the Seattle Police Department has become increasingly aware of privacy concerns as technologies like ALPR and body cameras take a more central role in day-to-day police work, they are legally obligated to provide information under the PRA. Attempting to protect privacy through noncompliance requires that the agency demonstrate that giving out the requested information would be highly offensive to a reasonable individual to disclose and that the information is of no legitimate concern to the public. Also, the burden of proof in demonstrating the potential violation is at the level of a tort, which in common law, is at the level of a preponderance of the evidence.[142] This means that the majority of evidence in a case must be in support of an agency's decision to deny a public records request for the case to stand in court. This level of proof is difficult and time-consuming to prove; demonstrating that privacy protection is warranted even when there is an applicable exemption can require highly trained legal staff and can place a large burden on limited agency resources.

[142] McKindles, John. "Differing Burdens of Proof in Contract Claims, Tort Claims." Last modified June 2010. Accessed May 14, 2016. http://www.mckindles-law.com/articles/burden-of-proof.htm.

Additionally, the PRA requires agencies and courts to make decisions about what records require privacy protection based solely on the attributes contained within the record itself. Agencies are not allowed to consider whether or not information could be joined or associated with information held in other databases to derive new, privacy-invasive information about individuals. They are only allowed to assess the privacy implications of the attributes included in the requested record itself. This means that the fact that license plate numbers can be used to join records from multiple sources including government records, insurance records, or background checks is irrelevant in the agency's decision to deliver the requested information in a public records request. This legal limitation on privacy regulation is at the core of the challenge to protect privacy through regulation and policy; the major findings of the scholarly literature about the privacy implications of big data almost all rely on the linkage argument, which is the idea that records from multiple sources can be joined together to derive new information about individuals. If sensor program managers and privacy advocates in municipal agencies are not allowed to use the linkage argument to justify privacy protection, they are essentially forced to fight with one hand tied behind their backs. In other words, in cases where data is not covered by a PRA exemption, the language of the PRA makes it almost impossible to mitigate the possible privacy implications of releasing the entirety of the dataset.

Unwarranted noncompliance is also financially risky for agencies. The penalty for noncompliance, if found to be unwarranted in court, can be assessed on a per-page basis,[143] which means that unanswered requests for large databases could cost an agency hundreds

[143] MRSC of Washington 2015.

of thousands of dollars. The penalty for noncompliance is steep because the PRA is supposed to ensure transparency and allow for citizen oversight of government operations, but it has the side effect of making it increasingly difficult to protect privacy.

## Open Data

The Seattle Police Department is part of the White House's Open Police Data initiative and works closely with the City of Seattle to prepare datasets of public interest for release to the open data portal hosted by the City of Seattle and Socrata, a third-party data warehouse management company. Every dataset that is released by the police department is vetted by the Seattle DoIT and is checked via a Privacy Impact Assessment. In most cases, the records that are released through the city's open data portal are either aggregated to at least a block-group level or have all identifying attributes removed. However, even if all PII is removed, data still retains spatial identifiers like latitude and longitude. Interviewees at SPD stated that they have no interest in publishing the ALPR data via the open data portal because it is not of substantial interest to the public. However, a public records request for four months of ALPR data was turned into a publicly available time series and map that display the data in much the same way that a government-sponsored online mapping tool would.[144] This is a clear example of ways in which PRA requests can generate the same result as publishing data to online data portals. However, because the Open Data Initiative is voluntary, the agencies have much greater discretion over what data should be made public and therefore can work to protect privacy more easily.

[144] Newell, Bryce. "ALPR Scans: One SPD Patrol Car over a Single Shift in 2013." Last modified October 23, 2015. Accessed May 14, 2016. https://public.tableau.com/profile/bcnewell#!/.

The Open Data initiative has had the counterintuitive effect of generating more PRA requests rather than reducing the overall number received. Interviewees at SPD stated that public records requests spike after they release a dataset, especially if that dataset has been redacted, aggregated, or otherwise altered to protect privacy. Interviewees stated that this was because people know that almost all data is accessible via a public records request, so they are not easily satisfied with an altered dataset. One interviewee stated that this was a direct result of the breadth of the PRA; in her experience, other states without incredibly broad public information laws do not experience the same increase in requests for records after the release of data through the White House Open Data Initiative.

This is an important piece of feedback because one of the major justifications for the expense of creating, managing, and hosting a data portal is that it reduces the overhead cost of answering PRA requests. If the Open Data Initiative actually causes PRA requests to increase rather than decrease, it undermines the economic justification for participating in the Open Data Initiative. However, the assertion made by interviewees at the SPD was not verified empirically or discussed during any of the other case study interview. Thus, it is not possible to say that publishing data to the open data portal universally causes public records requests to increase. This is an area of research that deserves further consideration in the future.

### Public Outreach

The nature of the ALPR system makes it difficult, if not impossible, to deliver notice of data collection at the time of data collection. Unlike the Acyclica sensor network, which could theoretically use privacy beacons to alert smartphone users to the presence of the sensors, ALPR data is not capable of broadcasting information. Thus, if the SPD were to

attempt to deliver notice at the time of data collection, it is likely that they would simply post signs on police cars equipped with ALPR technology. However, this does not solve the problem because the license plates can only be recorded when the police car is behind the vehicle in question. Thus, the driver of the vehicle would not be able to see the notice before data was collected.

There is also no opt-out capability with the ALPR system. This is a function of the system design, and could potentially be rectified if sufficient monetary and political forces supported the effort. However, the Seattle Privacy Policy only requires that an opt-out option be provided whenever *possible*, so the current ALPR system is not in violation of the City's privacy policy. Information about the ALPR system is hosted on the SPD's website. The information includes an overview of the program as well as the statutory language that allows the police department to use vehicle-mounted ALPR sensors.

### *ORCA Smart Cards*

The ORCA smartcard technology relies on RFID chips installed in each card. Card readers on buses or at transit stations read the cards and record the fare payment transaction. The data logs for each card reader are backed up nightly and sent to Vix's off-site server. Because every fare payment transaction record has financial information in it, ORCA data is not subject to the PRA. Thus, the software architecture of the system and the data release process is drastically different than it would be if the ORCA was subject to the PRA. The following sections will discuss the findings and implications of the ORCA smartcard case study.

## Data Release

As stated above, ORCA data is only released for explicit research projects or to large employers who have purchased ORCA passes for their employees as part of a commute reduction program. Because of initial negative feedback about privacy and data security concerns, Sound Transit and ORCA program managers have been very wary of giving out data. In fact, Sound Transit has only released the entire database once. After lengthy discussions with the Washington State Transportation Center (TRAC) and researchers at the University of Washington, Sound Transit agreed to release the data for research purposes. As a part of that application process, researchers completed an Internal Review Board (IRB) assessment, developed a research methodology, and wrote a privacy and data protection plan. Even with those safeguards in place, any change in methods or research questions requires an additional IRB assessment.

Researchers working with the ORCA card data are very familiar with the privacy literature and understand that multiple attributes can be used to identify people through locational analysis. They work closely with Sound Transit to ensure that any research project that they undertake does not lead to unintentional privacy violations for the community. Once again, this level of ethical consideration is only possible to enforce because Sound Transit is not required to release data; they have the freedom to ensure that people with access to the data are committed to ethical data management practices.

## Data Access Levels

One of the side effects of Sound Transit's reluctance to share data is that the public has less access to transit ridership data than is desirable. Although King County Metro provides user-friendly charts that display system-wide ridership statistics, route-specific

ridership data is not easily accessible. Ridership data is useful in judging the economic efficiency of transit, which is an issue that the region has hotly debated in recent years. ORCA data is an easy source of ridership data. It could be used to provide interested citizens and decision makers with a detailed picture of route-level ridership data. While ensuring that data is not used in ways that compromise individual privacy is important, there is no reason that the data should not be used to generate information about system usage trends. Such information is critical to making good policy decisions and is one of the many benefits of smart city sensor programs.

Because ORCA is not subject to the PRA, program managers have the unique opportunity to design an access architecture that differentiates levels of data access based on the role of the individual or organization requesting data. Under the PRA, all information requests are anonymous; nothing more than an email is required for information delivery. With the yet-to-be-created ORCA data portal, the level of data accessible could be determined by the role that the user fits in to. Agency officials imagine this as a three-tiered system with one set of data that provides summary statistics at a meaningful, but generalized level that is designed for the general public. Like Census data, anonymized and geographically aggregated records would be available to the general populace via an online data portal. Accessing the public data portal would not require any sort of identification or verification. The second level would be comprised of bus stop-level data that has been stripped of all identifying attributes and aggregated to provide a level of geographic cloaking for individuals who live in more remote areas and thus might be part of a less-populated dataset. This type of data would function much like Census data, where the size of the base geography is partially determined by population. The final

level of data access would be unrestricted access to raw ORCA data. To gain access to this level of data, the requester would have to pass verification tests, present a research plan, and agree to abide by a code of ethical conduct. This is precisely the approach already employed by the U.S. Census Bureau, which provisionally grants researchers access to individual-level data to facilitate research projects that have the potential to advance the common good.

The multi-tiered data access levels is a tested access method that has the dual benefits of allowing interested parties to access data and of protecting the privacy of individuals. This more nuanced approach to data management is far preferable over the all-or-nothing approach of the PRA or of Sound Transit's current data management methods, respectively. For the smart city to deliver on its promise of delivering more information about city services and of helping to drive change through the use of big data, program managers need to find a way to both protect privacy and allow data to be used. The three-tiered approach that Sound Transit is currently considering for ORCA data is one of the better approaches to managing sensitive information available to public agencies.

## Geographic Opacity

Interviewees stressed that ORCA data is unique among location-based data sources because there is already a high degree of obfuscation built into the data. Because data is collected when people board a bus or train and not when they leave their origin point, travelers' exact origins remain unknown. This is especially true in dense areas of the city that have many bus stops, a dense network of streets, and a large population. The geographic opacity of the data works to protect privacy in the busiest transit corridors and in the densest parts of the city, where thousands of people board buses on a daily basis. It

is less effective in smaller communities where ridership is lower, fewer people live within walking distance of the bus stop, and distances between bus stops are greater. Under those conditions, the geographic opacity provided by bus stop level data collection is not a privacy shield. This is one of the reasons that the size of the ridership base must be considered when determining what data can be published without compromising locational privacy.

In summary, ORCA program managers have the freedom to create a data sharing program that is structured around the unique qualities of the ORCA data because they are not bound to comply with the PRA. Only the exemption from the PRA makes it possible for the agency to require ethical reviews before distributing data. While this operating policy has important implications for government transparency, it is clear that the ORCA program's concerns about privacy have helped to create a program that attempts to keep transparency and privacy in balance.

# 5. Discussion

The review of the existing theoretical, legal, and empirical literatures has made it clear that privacy is a long-standing, complicated problem that is only exacerbated by new technologies that make it increasingly easy to collect highly detailed data. According to a 2014 report by the FTC, data brokers—companies that collect consumers' personal information and resell or share that information with others—rely on government data obtained through public records requests to build profiles on nearly every U.S. consumer.[145] This complex, secretive, and largely unregulated industry uses public records to enable targeted marketing specifically tied to each individual's interests and spending habits. The FTC report established that there is a strong market for government data that uses datasets in ways far removed from their intended purposes.

As municipalities continue to invest in smart city data collection technologies, it is important that decision makers and program managers consider the range of potential ramifications that could result from the collection of disaggregated, location-based data. Because city agencies both have a duty to protect the interest of the public and are bound to comply with the Freedom of Information Act and the Public Records Act, considering the implications of data collection is an absolute necessity. The following is a discussion of the three case studies and the implications of the use of smart city data collection programs in the context of the academic and empirical literature presented in Chapter 2.

### *Notice and Consent*

[145] United States. Federal Trade Commission, Author. *Data Brokers : A Call for Transparency and Accountability*. Washington, D.C.]: Federal Trade Commission, 2014.

While the literature regarding notice and consent have made it evident that privacy self-management through notice and consent regimes is a flawed mechanism, there is not a clear case for completely dismissing the principles themselves. Daniel Solove and others have demonstrated that there is much to be desired from notice and consent; notice and consent do not solve the problem of managing downstream use of data, nor does it address the problem of user apathy.[146] However, providing notice of data collection, is one of the most important transparency safeguards that can serve to limit needless data collection by municipal agencies. If people are informed of how their data is collected, used, and distributed, they will be better able to either request that the data be shared for public purpose or that the data collection be halted. This remains true even if the technology does not have an opt-out consent structure; political processes exist for concerned citizens to make their concerns known.

That is, municipal agencies have an ethical responsibility to operate in the interest of the public good. Data should only be collected for clearly defined public purposes, and information regarding what data is collected and how it is managed should be freely available to the public. If individuals feel strongly that the city should not be collected detailed information, cities should listen. This is the intent behind Nissenbaum's concept of informational norms; people who are giving up data should have a say in what they are relinquishing, who they are relinquishing it to, and what is done with that information. The

[146] Solove, Daniel J. "Privacy Self-Management and the Consent Dilemma." Harvard Law Review 126 (2013): 1879-2479.

only way that this can be achieved is through a committed and principled provision of notice for each data collection program.

### *Data Management through Internal Review*

ORCA, Sound Transit's regional transit pass, has used the fact that each record is linked to financial information to justify not releasing person-level trip data under the Public Records Act. Because the PRA has an exemption for any financial transaction data, the records of trips are protected from the level of public scrutiny that ALPR data is subject to. While this is potentially is a positive aspect that prevents architectural privacy harms from wholesale release of person-level datasets, it may also be possible that current data management practices limit transparency. Although this thesis is primarily about privacy risks, it is also important to balance privacy risks with the need for transparency and government oversight. Sound Transit has a privacy policy, but it does not have a public-facing transparency policy in the way that the City of Seattle does. This means that the agency makes decisions about releasing data on an ad-hoc basis.

Right now, Sound Transit has only released the entirety of its ORCA database to TRAC, the Washington State Transportation Center, a multi-university research group. They were required to submit data management plans, complete an IRB review, and outline research methodologies before they were given access to the data. While this has resulted in a secure data handling program, it is also fair to say that this level of authentication has made it difficult for other researchers to gain access to the data.[147] Because there is no coordinated data release plan for trip data, Sound Transit's options are

[147] For example, the author requested access to ORCA trip data for a graduate school project and was denied due to privacy concerns.

to release all data or to release no data. Ideally, Sound Transit would find a way to aggregate or sufficiently anonymize data and could release system ridership data for all transit agencies that support the ORCA smartcard. This could provide a new data source for innovation in transportation, and would definitely help transit planners assess the efficiencies of transportation agencies in the Puget Sound region.

Essentially, the best way for Sound Transit to manage its data is to have a tiered access structure. For the first level, they could create aggregate statistics and maps designed to provide a general overview of system performance. This data would be designed to help inform the general public and would require no data manipulation or processing by the end user. The second level would be an open data portal similar to the one operated by the City of Seattle, where they could publish data for public use. This data would be anonymized and aggregated to a level found to be sufficient to protect locational privacy, but still granular enough to provide insight to researchers and interested individuals. Civic hackers and other tech-oriented individuals could use this information to create new applications and programs, thereby fueling innovation and civic engagement. People using this data would agree to abide by the agency's data policies, but would not need to submit individualized research plans or complete an IRB review. The third level of data would be the raw trip data; data could be anonymized, but each individual trip would be stored as an independent record. To access this level of data, researchers would have to submit data management plans, methodologies, and complete an IRB review. However, the process would be formalized and made clear to the general public; data release decisions would not be made on an ad-hoc basis.

By using a tiered data release system, where different levels of information are available for different audiences, Sound Transit can both protect privacy and use the richness of their datasets to promote innovation, research, and growth. Because the fare transaction data is not subject to the PRA, Sound Transit has a unique opportunity to be a thought leader in the best way to balance transparency and privacy in data management policy.

### *Strengths and Weaknesses of Privacy by Design*

Privacy by design, the principle of designing data collection systems to avoid or minimize the collection of sensitive information, has the potential to transform the way in which data is collected, manipulated, and stored. It potentially can reduce the risk of architectural privacy harms and can also reduce the post-collection data cleaning necessary to prepare datasets for public release. If done correctly, privacy by design would allow cities to collect necessary information without gathering data that could be used to track or profile specific individuals. In some ways, privacy by design seems like a magic bullet; instead of having to sacrifice privacy for efficiency (or vice versa), cities can simply design systems that ignore or discard all personal information and focus solely on aggregate statistics relevant to service delivery.

However, the Acyclica case study shows that even in a highly advanced system, some personal information is likely to be collected. The Acyclica sensors collect and encrypt MAC addresses before creating aggregated corridor summaries. While only the aggregated summaries are sent to the City of Seattle, the hashed trip data is still retained by Acyclica in its private servers. This may not be an issue, but it demonstrates that even when privacy by design principles are followed, it is likely that the data collected will still be of a deeply

personal nature. That is, privacy by design principles are limited by the questions the sensor system is designed to answer. If person-level data is needed to answer a particular question, then even the use of privacy by design principles will not change the fact that it is still information that can be linked to individuals. Thus, it is important to recognize the limits of privacy by design principles, even as we begin to use them as a way to mitigate privacy risks and reduce unnecessary data cleaning work. Essentially, there is no easy answer to a problem as complicated as privacy; it is important to recognize that even in the best of circumstances, risk will still exist. We must weigh this reality against the potential benefits of data collection before creating new data collection programs.

### *Seattle's Attempt to go Beyond the PRA*

The City of Seattle is a thought-leader in the field of municipal data privacy. At the request of the Mayor and the City Council, city staff created a new privacy program, which includes a revamped privacy policy, informational materials about the privacy program, and a privacy toolkit designed to determine the levels of risk presented by datasets. The privacy program has helped the City create an open data portal that provides data for research and public information while still ensuring that the released data is not likely to be used to build individual-level profiles or to track people as they move through the city. Part of the reason that this effort has been so revolutionary in the redefining the City's role in protecting privacy is that the privacy policy states that:

> Personal information is any information relating to an identified or identifiable individual. Examples of personal information include but are not limited to a person's name, home or email address, social security number, religion, political opinions, financial and health records, location and racial and ethnic origin. [148]

[148] City of Seattle. "Privacy Program: Facts & Questions." Last modified June 2015. Accessed May 14, 2016. http://www.seattle.gov/Documents/Departments/InformationTechnology/privacy/PrivacyFactsLB.pdf.

This wide definition of privacy is indicative of the city's commitment to protecting privacy; they have actually gone beyond the stated levels of the policy and have refrained from collecting personally-identifiable internet traffic metadata and from requiring identity verification for use of most online city resources. The multi-disciplinary Privacy Advisory Committee helped the City craft a privacy policy that is capable of protecting publicly collected data.

However, the City's definition of privacy is much broader than the definition of Personally Identifiable Information contained in the PRA. The PRA defines personal information as:

> (1) Personal information in any files maintained for students in public schools, patients or clients of public institutions or public health agencies, or welfare recipients.
> (2) Personal information, including but not limited to, addresses, telephone numbers, personal electronic mail addresses, social security numbers, emergency contact and date of birth information for a participant in a public or nonprofit program serving or pertaining to children, adolescents, or students, including but not limited to early learning or child care services, parks and recreation programs, youth development programs, and after-school programs. Emergency contact information may be provided to appropriate authorities and medical personnel for the purpose of treating the individual during an emergency situation.[149]

This definition makes no mention of location, address, political identification, or ethnic identity. It is a much more limited definition of privacy, and focuses solely on PII. This has deep ramifications for locational privacy because the PRA requires that all records associated with a government function be disclosed unless the data in question is either included in the definition above or explicitly covered by an enumerated exemption in the

[149] MRSC of Washington. "Public Records Act for Washington Cities, Counties, and Special Purpose Districts." Last modified 2016. Accessed May 14, 2016. http://mrsc.org/getmedia/796a2402-9ad4-4bde-a221-0d6814ef6edc/publicrecordsact.pdf.aspx?ext=.pdf

law is exempt from disclosure requests. Exemptions include investigation details, human resource records for government agencies, real estate appraisals, financial information, library records, utility metering, transit ridership, traffic speeds, and vehicle location information, health care data, and trade secrets. While this is not an exhaustive list of exemptions, it serves to demonstrate that there are few exemptions that can help to protect data collected through smart city sensor systems. Although ORCA data are protected because they are records of fare payment transactions, the other sensor systems studied do not qualify for the same protection. As was seen in the review of the case study systems, whether or not the system is subject to the PRA played a deterministic role in the design and structure of the data management system used in each case study.

The Washington State legislature has recognized that the limited range of exemptions and the thematic breadth of the PRA was not necessarily in the best interest of the public given new technologies and data management techniques. In 2007, the legislature passed Senate Bill 5437, which established the Sunshine Committee and tasked it with making recommendations for new or revised PRA exemptions based on public input and expert opinion. It has been an active part of the privacy advocacy infrastructure in Washington State, but it tends to focus on more individual-focused privacy issues (e.g. protecting the identity of children who have been the victims of sexual abuse) rather than addressing the larger conceptual problems that make the distribution of certain types of public records deeply problematic from the perspective of privacy protection.[150] The reasons for this

[150] Brown, Melinda. "Sunshine Committee." Washington State Office of the Attorney General. Last modified May 2016. Accessed May 14, 2016. http://www.atg.wa.gov/sunshine-committee.

focus are not entirely clear; perhaps this is a result of political pressure, or a result of the structure of the law. However, we must reimagine the ways in which the PRA allows access to government records if the PRA is to both ensure government transparency and provide adequate privacy protection given the increasingly complex informational and technological environment.

### *Reimagining the Public Records Act*

To reshape privacy protection in Washington State, we must find a way to change the Public Records Act to reflect the realities of modern information technology and data management. This includes acknowledging both the linkage argument (the fact that big data relies on drawing inferences from multiple sources of data) and the fact that there is a fundamental difference between the inferential potentials of a single record and a complete database. The release of full databases creates the potential for architectural harms in a way that the release of single records do not. Architectural harms are informational privacy violations that are similar to environmental degradation or pollution; even if a specific instance of a violation does not seem problematic in isolation, when all potential harms are taken in aggregate, the enormity of the harm becomes clear.[151] The fact that the PRA allows for the dissemination of complete databases poses a challenge for privacy program managers.

We must also create exemptions to protect information that can easily be used to merge records from multiple databases like license plate information, and address records. Essentially, I am arguing that municipal privacy protection in Washington State cannot be complete until the PRA is updated. Below I suggest some potential revisions for the PRA

[151] Solove 2006, 8.

that incorporate the ideas presented above. While these proposed changes would need to be vetted by a legal team before they were adopted, they offer a direction for positive change in Washington's Public Record Act.

### *Changes in Language in the PRA*

Text in italics denotes proposed language.

#### Requirements

*(1) Records requests may not be generated through the use of an automation program. A captcha may be used to verify the principle of non-automation.*

*(2) If an entire database is requested, the requester must prove that he or she has a legitimate interest in the data, which may be, but is not limited to, research for public purpose.*

*(3) Individuals or corporations who receive data from municipal agencies must comply with the privacy policy of the agency that originated the data.*

#### Definition

(1) Personal information in any files maintained for students in public schools, patients or clients of public institutions or public health agencies, or welfare recipients.

(2) Personal information, including but not limited to, addresses, telephone numbers, personal electronic mail addresses, social security numbers, *vehicle license plate information, location information including but not limited to home address, records of location*, emergency contact and date of birth information for a participant in a public or nonprofit program serving or pertaining to children, adolescents, or students, including but not limited to early learning or child care services, parks and recreation programs, youth development programs, and after-school programs. Emergency contact information may be provided to appropriate authorities and medical personnel for the purpose of treating the individual during an emergency situation.[152]

#### Exemptions

*(1) Privacy managers at city agencies, shall, in accordance with best professional judgment, make determinations regarding the threat of linkage and or inference between datasets. If there is a clear or present threat that distribution of the requested data could be used to infer private information, the privacy manager shall redact the information in question and shall determine whether to release the data to the requester.*

[152] MRSC of Washington 2016, 6.

## Contributions to the Field

While it is clear that much research remains to be done on the subject of smart city data collection programs, this thesis brought several important aspects into sharper focus and provided fodder both for policy change and for future research. After reviewing the literature surrounding privacy by design, privacy law, the ethics of privacy, and local privacy policy, it is clear that a right to privacy in public needs to move from the theoretical literature to the realm of policy and law. Right now, the FTC, the main federal agency responsible for regulating privacy lacks strong regulatory tools and has relied on ancillary issues to extend its authority to privacy issues. In Washington State, the PRA has effectively crippled efforts to protect privacy by requiring that records not expressly exempt from disclosure be delivered in their entirety upon request. This research effort has examined the data collection, management, and retention practices of three Seattle-area data collection programs. It has demonstrated that the Washington State Public Records Act has dramatically shaped the design and implementation of smart city sensor networks.

For technological reasons, each system is designed and managed in different ways, but the high-level difference between the systems is that some of the data is subject to the PRA while other datasets are not. In the systems where all data collected must be disclosed through PRA, the only recourse agencies have to protect individual privacy is to create systems where the city never collects disaggregated data. For example, the use of these privacy by design principles ensures that even if data is requested under the PRA, there is no cause for concern about individual privacy, as long as the contract between Acyclica and the City of Seattle prevents selling data in secondary markets. The Acyclica sensor network was able to apply privacy by design principles to ensure that data on traffic flows

could be collected through the use of hashed trip identifiers. While this is still not an ideal solution (hashed identifiers can still demonstrate unique travel patterns), it is better than if the system were to use unencrypted MAC Addresses as a basis for data analysis. The ALPR system has yet to have the same level of success in reimagining the data collection process and thus still poses a challenge for locational privacy protection. The ORCA data collection program is in the best position to protect privacy because it is not subject to the PRA. Thus, it is free to create its own privacy-protective data management practices and to implement best practice approaches for handling sensitive data.

In essence, my research revealed that the Public Records Act has a profound effect on data handling and privacy protection policies at municipal agencies in the City of Seattle. It also revealed that privacy can be protected through tiered release programs, like the one the Sound Transit is considering. While both approaches have merit, advocates must first look to altering the Public Records Act to reflect the realities of the big data era if privacy protection is to be a priority in Washington State.

### *Future Directions*

While the fact remains that much of the proposed language above may not become law, it may be possible to begin to present the ideas and best practices discussed in this paper to the Public Records Exemption Committee. The committee does not explicitly solicit new exemptions from interested parties, it does accept public comments on exemptions currently being considered. I believe that with some behind-the-scenes work with committee staff, it might be possible to present the results of this research project for consideration and potential inclusion the Public Records Act. If that could be done,

municipal agencies in Seattle and elsewhere in Washington State would be better able to operate transparently while still offering a high level of privacy protection to the citizens.

In the short term, it may be better to actively pursue privacy by design solutions. As the Acyclica case study demonstrated, privacy by design can be used to eliminate privacy risks before data is ever analyzed by city staff. Although the sensor network records and encrypts MAC addresses, the trip data through corridors is aggregated into 5 minute intervals reflecting average speed through arterial corridors. When data is sent to the city, it is already aggregated, so no individual locational information can be inferred from the city's version of the data.[153] While privacy by design principles must be recalibrated to every data collection program, it is important to remember that data minimization (the process of collecting only data necessary to answer a defined research question) and privacy by design principles can be used to create sensor networks that achieve their intended goals without compromising individual privacy. Future research should focus on finding new technical solutions to these types of problems, and should seek to demonstrate privacy risks of each system design before implementation at the city-scale.

On an academic level, there is clearly more work to be done in the field of municipal privacy. For one, work remains to be done to demonstrate exactly how architectural privacy harms affect both individuals and society. Right now, much of the literature focuses on anecdotes of data breaches or descriptions of the practices of data brokers. While these are important first steps, much remains to be done to demonstrate that privacy

[153] Of course, this begs the question of whether or not a third party contractor like Acyclica should be managing municipal data streams. While this is an important question that has important privacy implications for cities and their residents, it is not within the scope of this thesis to attempt to provide a compelling answer to this problem.

harms have deep and systemic impacts on personal freedoms and on economic outcomes. Work also remains to be done on how Nissenbaum's theories of informational norms relate to Solove's taxonomy of privacy. Questions remain to be answered about consent and the shifting nature of societal norms. The answers to these questions are not yet clear, so it is difficult to articulate a compelling justification for strong privacy protection when confronted with more easily articulated justifications for freedom of speed and governmental transparency.

While academic research, system design, and policy revision are all crucial, it is also important that planners and planning students begin to think about their role in the future of the smart city. Networked sensor technologies generate planning datasets and will continue to be a crucial part of city service provision. Planners are the people in a municipal government who should be looking into the future to decide what infrastructure systems should be funded and what issues should be studied to increase accessibility, equity, and affordability. The fact that strategic plans and comprehensive plans do not have a dedicated section for ICT infrastructure is a glaring oversight that should be rectified as quickly as possible. Planners can influence privacy by setting goals and visions for the future of ICT infrastructure in their cities, and can bring those visions to life by reserving funding for ICT systems that comply with privacy guidelines laid down in the city's comprehensive plan. This basic toolset of planners is perfectly designed to begin to affect change; not using it is an egregious oversight.

To prepare for their role as future thought-leaders in city government, planning students should immerse themselves in data analytics and infrastructure management. Right now, many planning programs focus primarily on traditional planning activities.

There is a resistance to the idea that the planning field can include cutting-edge technologies; personal experience shows that interest in new technologies is met with confusion and indifference. Young planners are the industry's best chance of redefining itself; we need to embrace the role of technology in the future of our cities rather than limiting our studies to transportation, land use, and the myriad other specialties that we are comfortable planning for. We have a responsibility to look to the future; we cannot do that by closing our eyes to the change right in front of us.

Clearly, locational privacy and smart city sensor networks are developing fields. Articulating justifications, impacts, remedies, and protections for locational privacy has been the primary focus of this thesis, but the theoretical work and empirical work will need to adapt and evolve as technologies adapt and evolve. There is room for growth in the field; both empirical and theoretical work is needed to match the fast pace of growth in technological abilities.

## Limitations and Constraints

Although this work touches on many of the issues related to locational privacy protection in municipal data collection programs, it is not exhaustive. There is much that it did not touch on or that it did not examine because the issue of privacy is too large to fully examine in one research project. In the space and time available, only issues and programs somehow related to urban planning and locational privacy were considered. This means that issues like malicious attacks on databases, the effects of data brokers on the information economy, and the transaction costs associated with data management and privacy protection were not considered. In the space and time available, only a small

number of smart city sensor networks were considered and a limited number of program managers were interviewed. Interviewees with decision makers, privacy lawyers, and advocates working in the field would have made for a more well-rounded perspective.

Finally, because ICT is a fast-moving field, and because the policy environment discussed is unique to the City of Seattle and Sound Transit, this research does not represent a complete review of all of the ways that privacy can be protected through an amalgamation of privacy by design, notice and consent, and law. Instead, it presents a picture of three distinct systems operating in similar policy frameworks with startlingly different responses to the requirements of the Open Data Initiative, the Public Records Act, and the City of Seattle's Privacy Policy. These three case studies may represent a typology of data management techniques used in the City of Seattle and at Sound Transit, but further research is needed to fully understand data management practices in municipal agencies.

# 6. Conclusion

Sensor networks are now a key part of the information infrastructure used in municipal agencies. They provide a data collection service and allow city agencies to provide efficient service at the level people have come to expect. There is no easy way to replace sensor networks in the information architecture of city agencies, so these systems will continue to be used regardless of whether or not they pose a threat to informational privacy. Thus, urban planners and decision makers must work to mitigate privacy threats from smart city sensor networks by employing privacy by design principles, by delivering notice and providing an opt-out option where possible, and by crafting law and policy to balance the dual interests of privacy and transparency.

Privacy concerns are not new. Since technology made it possible to record information, people have been concerned with protecting privacy. In the past, privacy concerns were linked more closely to property or to the sanctity of the home because privacy was equated with the right to not be seen, but modern privacy scholarship has linked privacy with the rights to self-discovery and autonomy. Given the widespread adoption of sensor technologies and data collection programs, it is critical that theorists, lawmakers, and policy makers rethink privacy and acknowledge that a more robust understanding of privacy is needed in light of unprecedented technological developments. While privacy scholars have made these concerns clear, it is evident that the laws and policies that control data management protocols at municipal agencies are only beginning to adapt.

The challenges of protecting privacy through policy and law were clear in all three case studies. Because the Public Records Act applies to all information related to the

provision of a public service and only offers a limited number of exemptions, few of which are relevant to protecting privacy of entire datasets, attempts to protect privacy at the agency or city level are curtailed. Using privacy by design principles to avoid collecting disaggregated data or using third-party contractors to collect data and compute summary statistics have been the two main workarounds to avoid releasing highly sensitive data through the Public Records Act. However, these workarounds are unsustainable. To give personal privacy the level of protection that it deserves, we need to rethink our legislation and work to make significant changes at the highest level of Washington State government.

While all three of the case studies considered in this research project were agencies in the City of Seattle or Sound Transit, this thesis still has important implications for program managers concerned with locational privacy in other parts of the United States. Firstly, it shows that privacy concerns are mostly managed at the state level, meaning that levels of data protection are likely to vary widely from state to state. Because there is no federal law that governs locational privacy, each state is individually responsible for protecting locational privacy. Secondly, it shows that when laws and policies are not designed to protect privacy, it may still be possible to use privacy by design principles to design sensor networks that only store aggregated information. This may be the best solution in many cases because it limits the potential for data breach and helps ensure that privacy is not compromised through public records requests. Finally, this research is important because it demonstrates that privacy policies are most successful when they are implemented proactively rather than reactively. This is crucial for the implementation of tiered data release systems and for

privacy by design solutions. Mechanisms for data release must be visualized before the system is implemented for the data release effort to be both privacy-secure and efficient. For example, the ALPR system was implemented before locational privacy was a major concern, so the system designers did not design the system with privacy in mind. If the system was rebuilt today, there is no way that unencrypted scans of license plates, locations, and positive matches with wanted persons records would be integral parts of the system. That is to say, by thinking about privacy *before* implementing data collection systems, agencies can avoid costly back-end aggregation or anonymization practices, and can also avoid the high labor costs associated with large-scale data requests.

In conclusion, locational privacy will become more of a concern in coming years; protecting privacy while still embracing the potential of the new technologies will require both legislative action and the use of privacy by design principles. This issue will be one of the critical factors in the success or failure of smart city sensor networks—if agencies are unable to find ways to both protect privacy and collect data, they may in fact choose to not use these technologies. Thus, the time for action is now; we have one chance to create the privacy-aware smart city.

# Bibliography

Allamaraju, Kingravi, Axelrod, Chowdhary, Grande, How, Crick, and Weihua Sheng. "Human Aware UAS Path Planning in Urban Environments Using Nonstationary MDPs." Robotics and Automation (ICRA), 2014 IEEE International Conference on, 2014, 1161-167.

Angelidou, Margarita. "Smart City Policies: A Spatial Approach." *Cities* 41 (2014): S3-S11.

Apple, Inc. "A Message to Our Customers." Apple. Last modified February 16, 2016. Accessed March 22, 2016. http://www.apple.com/customer-letter/.

Arendt, Hannah. The Human Condition. Charles R. Walgreen Foundation Lectures. Chicago: University of Chicago Press, 1958.

Armbruster, Ginger. "Cloud Computing: Considerations for Municipal Government." Master's thesis, University of Washington, 2013.

Armbruster, G., Whittington, J., & Endicott-Popovsky, B. (2013). “Threats to municipal information systems posed by aging infrastructure.” International Journal of Critical Infrastructure Protection, International Journal of Critical Infrastructure Protection, 2013.

Ashok, Ashwin, Viet Nguyen, Marco Gruteser, Narayan Mandayam, Wenjia Yuan, and Kristin Dana. "Do Not Share! Invisible Light Beacons for Signaling Preferences to Privacy-Respecting Cameras." Semantic Scholar. Last modified September 7, 2014. Accessed March 22, 2016. https://pdfs.semanticscholar.org/9f8d/cd4a1f08c430702c05e0c49a7ad411cba38c.pdf.

Barocas, Solon, and Helen Nissenbaum. “Big Data’s End Run around Anonymity and Consent." In *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, 44-75. Cambridge University Press, 2013.

Barocas, Solon, and Helen Nissenbaum. "Computing Ethics: Big Data's End Run Around Procedural Privacy Protections."*Association for Computing Machinery. Communications of the ACM* 57, no. 11 (2014): 31.

Batty, Axhausen, Giannotti, Pozdnoukhov, Bazzani, Wachowicz, Ouzounis, and Portugali. "Smart Cities of the Future." *The European Physical Journal Special Topics* 214, no. 1 (2012): 481-518.

Bertino, Elisa L., Michael Gertz, Bhavani Thuraisingham, and Maria Luisa Damiani. "Security and Privacy for Geospatial Data: Concepts and Research Directions." *Proceedings of the SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy in GIS and LBS, SPRINGL'08*, 2008, 6-19.

Bourmpos, Michail, Apostolos Argyris, and Dimitris Syvridis. "Smart City Surveillance Through Low-Cost Fiber Sensors in Metropolitan Optical Networks." *Fiber and Integrated Optics* 33, no. 3 (2014): 205-23.

Brown, Melinda. "Sunshine Committee." Washington State Office of the Attorney General. Last modified May 2016. Accessed May 14, 2016. http://www.atg.wa.gov/sunshine-committee.

Bureau of Information and Telecommunications. "Quarterly Newsletter: Winter 2013." South Dakota Bureau of Information and Telecommunications. Last modified February 2013. Accessed December 4,2015. http://bit.sd.gov/newsletter/2013/BIT%20Newsletter_February2013.htm.

Calo, M. Ryan. "Against Notice Skepticism in Privacy (and Elsewhere)." *Notre Dame Law Review* 87, no. 3 (2012): 1027.

Calo, M. Ryan. "The Boundaries of Privacy Harm." *Indiana Law Journal* 86, no. 3 (2011): 1131-1162.

Carnegie Mellon University. "Projects Library." MetroLab Network. Accessed March 22, 2016. http://metrolab.heinz.cmu.edu/index.php/projects/.

Catlett, Charlie, and Von Welch. "Urban Sensor Data Privacy Issues: Findings of the Array of Things (AoT) Privacy Breakout Group." *STREAM2015*, 2015.

City of Seattle. "City of Seattle Privacy Policy." Last modified April 2016. Accessed June 6, 2016. http://www.seattle.gov/Documents/Departments/InformationTechnology/privacy/PrivacyPolicyFINAL.pdf.

City of Seattle. "Privacy Facts." Last modified March 2016. Accessed June 6, 2016. http://www.seattle.gov/Documents/Departments/InformationTechnology/privacy/PrivacyFactsLB.pdf.

City of Seattle. "Privacy Program: Facts & Questions." Last modified June 2015. Accessed May 14, 2016. http://www.seattle.gov/Documents/Departments/InformationTechnology/privacy/PrivacyFactsLB.pdf.

Cottrill, Caitlin D., and Piyushimita 'Vonu' Thakuriah. "Privacy in Context: An Evaluation of Policy-based Approaches to Location Privacy Protection." *International Journal of Law and Information Technology* 22, no. 2 (2014): 178-207.

David Olsen, Jeffrey Johnson, Matthew Hacking, and Nicole Forsgren. "MUNICIPAL INFORMATION SYSTEMS: CURRENT PRACTICES AND ISSUES." Issues in Information Systems 2 (2001): 350-56.
Davis, K., & Patterson, Doug. (2012). Ethics of big data. Sebastopol, CA: O'Reilly.

Department of Information Technology. "City of Seattle Privacy Statement." City of Seattle. Last modified December 2014. Accessed December 4, 2015. http://www.seattle.gov/information-technology/privacy-program/privacy-statement.

DeCew, Judith Wagner. In Pursuit of Privacy : Law, Ethics, and the Rise of Technology. Ithaca: Cornell University Press, 1997.

De Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. "Unique in the Crowd: The Privacy Bounds of Human Mobility." *Scientific Reports*3 (2013): Scientific Reports, 2013, Vol.3.

Dryer, Randy L., and Stroud, S. Shane. "Automatic License Plate Readers: An Effective Law Enforcement Tool or Big Brother's Latest Instrument of Mass Surveillance? Some Suggestions for Legislative Action." *Jurimetrics Journal of Law, Science and Technology* 55, no. 2 (2015): 225.

Finch, Kelsey and Omer Tene. "Welcome to the Metropticon: Protecting Privacy in a Hyperconnected Town." Fordham Urban Law Journal 41 (2014): 1581-759.

Fritzler, Alan. "An Ethical Checklist for Data Science." Data Science for Social Good. Last modified September 18, 2015. Accessed December 5, 2015. http://dssg.io/2015/09/18/ethics-checklist.html.

Gellersen, Hans-W., Want, Roy, Schmidt, Albrecht, and LINK. *Pervasive Computing Third International Conference, PERVASIVE 2005, Munich, Germany, May 8-13, 2005 : Proceedings*. Lecture Notes in Computer Science ; 3468. Berlin: Springer, 2005.

Gutmann, Myron, P. Witkowski, Kristine Colyer, Corey O'Rourke, and JoAnne McNally. "Providing Spatial Data for Secondary Analysis: Issues and Current Practices Relating to Confidentiality." *Population Research and Policy Review* 27, no. 6 (2008): 639-65.

Heeney, Catherine. "Breaching the Contract? Privacy and the UK Census." *The Information Society*28, no. 5 (2012): 316-28.

Hill, Kashmir. "E-Z Passes Get Read All Over New York Not Just at Toll Booths." Forbes. Last modified September 2013. Accessed June 6, 2016. http://www.forbes.com/sites/kashmirhill/2013/09/12/e-zpasses-get-read-all-over-new-york-not-just-at-toll-booths/#5e2b61ef3cfc.

Hoofnagle, Chris Jay. *Federal Trade Commission Privacy Law and Policy*. New York: Cambridge University Press, 2016.

Hoofnagle, Chris Jay, & Whittington, Jan. (2014). Free: Accounting for the costs of the Internet's most popular price. UCLA Law Review, 61(3), 606-670.

Information and Privacy Commissioner of Ontario. Transparency, Privacy, and the Internet: A Municipal Balancing Act. Ontario: City of Ontario, 2014.

Jain, Dharmesh. "A Discussion of Spatial Data Privacy Issues and Approaches to Building Privacy Protection in Geographic Information Systems." *Assessment Journal* 10, no. 1 (Winter2003 2003): 5. *Academic Search Complete*, EBSCO*host* (accessed February 16, 2016).

Jefferson, R. (2010). Computer Science, Technology and Applications : Biometrics, Privacy, Progress and Government (Computer science, technology and applications). New York, NY, USA: Nova Science.

Khoshgozaran, Ali, Cyrus Shahabi, and Houtan Shirani-Mehr. "Location Privacy: Going beyond K-anonymity, Cloaking and Anonymizers." *Knowledge and Information Systems* 26, no. 3 (2011): 435-65.

Kim, Joon. "Making Smart Cities Work in the Face of Conflicts: Lessons from Practitioners of South Korea's U-City Projects." *The Town Planning Review* 86, no. 5 (2015): 561-85.

Konings, Bastian, Schaub, Florian, Weber, Michael, Mattern, Friedemann, Santini, Silvia, Canny, John, F., Langheinrich, Marc, and Rekimoto, Jun. "Prifi Beacons: Piggybacking Privacy Implications on Wifi Beacons." *Pervasive and Ubiquitous Computing Adjunct Publication Proceedings of the 2013 ACM Conference*, 2013, 83-86.

Kroman, David. "Seattle installs new system to track individual drivers." Crosscut. Last modified September 8, 2015. Accessed March 22, 2016. http://crosscut.com/2015/09/seattles-new-technology-tracks-how-we-drive/.

Lane, Julia I. *Privacy, Big Data, and the Public Good: Frameworks for Engagement*. 2014.

Martin, Kirsten. "Transaction costs, privacy, and trust: The laudable goals and ultimate failure of notice and choice to respect privacy online" First Monday [Online], Volume 18 Number 12 (15 December 2013).

Maxwell, Terry. "Toward a model of information policy analysis: Speech as an illustrative example" First Monday[Online], Volume 8 Number 6 (2 June 2003).

Mazumdar, Soumya, Paul Konings, Michael Hewett, Nasser Bagheri, Ian McRae, and Peter Del Fante. "Protecting the Privacy of Individual General Practice Patient Electronic Records for Geospatial Epidemiology Research." *Australian and New Zealand Journal of Public Health* 38, no. 6 (2014): 548-52.

McDonagh, Maeve. 2009. "The protection of personal information in public registers: the case of urban planning information in Ireland." Information & Communications Technology Law 18, no. 1: 19-38. Communication & Mass Media Complete, EBSCOhost (accessed December 2, 2015).

McDonald, Aleecia M. and Lorrie Faith Cranor. "The Cost of Reading Privacy Policies." I/S: A Journal of Law and Policy for the Information Society 4 (2009): 543-897.

McKindles, John. "Differing Burdens of Proof in Contract Claims, Tort Claims." Last modified June 2010. Accessed May 14, 2016. http://www.mckindles-law.com/articles/burden-of-proof.htm.

Moore, Adam D. Privacy Rights Moral and Legal Foundations. University Park, Pa.: Pennsylvania State University Press, 2010.

Moore, Adam, ed. Privacy, Security, and Accountability: Ethics, Law, and Policy. N.p.: Rowman & Littlefield International, 2015.

MRSC of Washington. "Public Records Act." MRSC. Last modified June 1, 2015. Accessed April 23, 2016. http://mrsc.org/Home/Explore-Topics/Legal/Open-Government/Public-Records-Act.aspx.

MRSC of Washington. "Public Records Act for Washington Cities, Counties, and Special Purpose Districts." Last modified 2016. Accessed May 14, 2016. http://mrsc.org/getmedia/796a2402-9ad4-4bde-a221-0d6814ef6edc/publicrecordsact.pdf.aspx?ext=.pdf.

Newcombe, Tod. "4 Tech Trends Changing How Cities Operate." Governing: The States and Localities. Last modified December 8, 2014. Accessed December 4, 2015. http://www.governing.com/columns/tech-talk/gov-technology-trends-local-government.html.

Newell, Bryce. "ALPR Scans: One SPD Patrol Car over a Single Shift in 2013." Last modified October 23, 2015. Accessed May 14, 2016. https://public.tableau.com/profile/bcnewell#!/.

Newell, Bryce Clayton. "LOCAL LAW ENFORCEMENT JUMPS ON THE BIG DATA BANDWAGON: AUTOMATED LICENSE PLATE RECOGNITION SYSTEMS,

INFORMATION PRIVACY, AND ACCESS TO GOVERNMENT INFORMATION." Maine Law Review 66 (2014): 397-591.

Nissenbaum, Helen. "Privacy as Contextual Integrity.(Symposium: Technology, Values, and the Justice System)." *Washington Law Review* 79, no. 1 (2004): 119-157.

"NYU CUSP Unveils First-Of-Its-Kind 'Urban Observatory' In Downtown Brooklyn." Center for Urban Science and Progress. Last modified October 28, 2014. Accessed March 22, 2016. http://cusp.nyu.edu/press-release/nyu-cusp-unveils-first-kind-urban-observatory-downtown-brooklyn/.

Office of the Press Secretary. "Obama Administration Announces New smart city Initiatives." The White House. Last modified September 14, 2015. Accessed March 22, 2016. https://www.whitehouse.gov/the-press-office/2015/09/14/fact-sheet-administration-announces-new-smart-cities-initiative-help.

"Open Data Portal." Last modified March 18, 2016. Accessed March 22, 2016. http://www.data.gov/.

Pew Research Center. "American's Attitudes About Privacy, Security and Surveillance." Pew Research Center. Last modified March 20, 2015. Accessed March 22, 2016. http://www.pewinternet.org/files/2015/05/Privacy-and-Security-Attitudes-5.19.15_FINAL.pdf.

Practical Law. "Data Protection in United States: Overview." Practical Law. Last modified July 1, 2015. Accessed December 4, 2015. http://us.practicallaw.com/6-502-0467.

Rajagopalan, Ramesh and Varshney, Pramod K., "Data aggregation techniques in sensor networks: A survey" (2006). Electrical Engineering and Computer Science. Paper 22.

Regan, Priscilla M. Legislating Privacy : Technology, Social Values, and Public Policy. Chapel Hill: University of North Carolina Press, 1995.

Rossi, Luca, James Walker, and Mirco Musolesi. "Spatio-Temporal Techniques for User Identification by Means of GPS Mobility Data." 2015.

Roy, Jeffrey. 2014. "Open Data and Open Governance in Canada: A Critical Examination of New Opportunities and Old Tensions." Future Internet 6, no. 3: 414-432.

"S. 2270: Location Privacy Protection Act of 2015." Govtrack.us. Last modified November 10, 2015. https://www.govtrack.us/congress/bills/114/s2270.

Scassa, Teresa. 2014. "Privacy and Open Government." Future Internet 6, no. 2: 397-413.

Schaub, Florian, Rebecca Balebako, Adam Durity, and Lorrie Faith Cranor. "A Design Space for Effective Privacy Notices." *Symposium on Usable Privacy and Security*, July 22, 2015. Schulz, Jerry. "Information Technology in Local Government." ICMA. Last modified October 31, 2015.Accessed December 4, 2015.

Scipioni, Marcello, Paolo, Dey, Anind, K., Chu, Hao-Hua, and Hayes, Gillian. "A Privacy-by-design Approach to Location Sharing." *Ubiquitous Computing Proceedings of the 2012 ACM Conference*, 2012, 580-83.

Seattle Police Department. "16.170 - Automatic License Plate Readers." Seattle Police Department Manual. Last modified June 2012. Accessed March 22, 2016. http://www.seattle.gov/police-manual/title-16---patrol-operations/16170---automatic-license-plate-readers.

Sempra Energy Utility. "Applying privacy by design." Last modified June 2012. Accessed March 22, 2016. http://www.sdge.com/privacypaper.

Shvartzshnaider, Yan, Schrasing Tong, Thomas Wies, Paula Kift, Helen Nissenbaum, Lakshminarayanan Subramanian, and Prateek Mittal. "Crowdsourcing Verifiable Contextual Integrity Norms." 2016.

Solove, Daniel J. "Privacy Self-Management and the Consent Dilemma." *Harvard Law Review* 126 (2013): 1879-2479.

Sound Transit. "ORCA Privacy Statement." Last modified May 14, 2012. Accessed March 22, 2016. https://www.orcacard.biz/ERG-Seattle-Institution/jsp/static/ORCA%20Privacy%20Statement.pdf.

Swedburg, Claire. "ORCA Puts Ferries, Buses and Trains on One Ticket." RFID Journal. Last modified October 22, 2009. Accessed March 22, 2016. http://www.rfidjournal.com/articles/view?5320.

Tellis, Winston. "http://www.nova.edu/ssss/QR/QR3-3/tellis2.html." The Qualitative Report. Last modified September 1997. Accessed March 22, 2016. http://www.nova.edu/ssss/QR/QR3-3/tellis2.html.

United States. Federal Trade Commission, Author. *Data Brokers : A Call for Transparency and Accountability*. Washington, D.C.]: Federal Trade Commission, 2014.

U.S. Department of State. "The Freedom of Information Act." U.S. Department of State. Last modified January 11, 2016. Accessed March 22, 2016. https://foia.state.gov/Learn/FOIA.aspx.

Van den Hoven, Jeroen, Martijn Blaauw, Wolter Pieters, and Martijn Warnier. "Privacy and Information Technology." In Stanford Encyclopedia of Philosophy. Palo Alto, CA: Stanford University, 2014.

Wall, Matthew. "Is facial recognition tech really a threat to privacy?" BBC Technology News. Last modified June 19, 2015. Accessed December 4, 2015. http://www.bbc.com/news/technology-33199275.

Whittington, Jan, Ryan Calo, Mike Simon, Jesse Woo, Meg Young, and Peter Schmiedeskamp. "Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government." *Berkeley Technology Law Review* 30, no. 3 (May 4, 2015): 1899-1966.

Wood, Colin. "Seattle Begins Three Year IT Consolidation." Digital Communities: For City and County Leaders. Last modified November 30, 2015. http://www.govtech.com/dc/articles/Seattle-Begins-Three-Year-IT-Consolidation.html.

Yin, Robert K. *Case Study Research : Design and Methods*. 3rd ed. Applied Social Research Methods Series ; v. 5. Thousand Oaks, Calif.: Sage Publications, 2003.

Zimmer, M. (2010). “But the data is already public”: On the ethics of research in Facebook. Ethics and Information Technology, 12(4), 313-325.

# Acknowledgements

I would like to take a moment and acknowledge the gracious support from my friends, family, and academic mentors. Jan Whittington, thank you for encouraging me to pursue my interests and for asking me to think broadly about urban planning as an inclusive discipline. Mark Purcell, thank you for agreeing to be on my committee even as my topic shifted. Your input and advice were invaluable at the end. Anna Seivert and Lizzie Moll, this thesis owes its existence to your unflagging support. Thank you for listening to me talk about privacy and data for hours on end; you kept me sane. Meagan Scott, thank you for reading early drafts of my work and for living with me even as books and discarded drafts slowly covered every surface in our apartment. And to my family, thank you for long years of unconditional support. Truly, I stand on your shoulders.