

Data Flow in the Smart City: Open Data Versus the Commons



Richard Beckwith, John Sherry and David Prendergast

Abstract Much of the recent excitement around data, especially ‘Big Data,’ focuses on the potential commercial or economic value of data. How that data will affect people isn’t much discussed. People know that smart cities will deploy Internet-based monitoring and that flows of the collected data promise to produce new values. Less considered is that smart cities will be sites of new forms of citizen action—enabled by an ‘economy’ of data that will lead to new methods of collectivization, accountability, and control which, themselves, can provide both positive and negative values to the citizenry. Therefore, smart city design needs to consider not just measurement and publication of data but also the implications of city-wide deployment, data openness, and the possibility of unintended consequences if data leave the city.

Keywords Open data · The commons · Data stewardship

1 Introduction

This paper explores the complex relationship between cities and data or, more accurately, the way that the citizens of a city want data about their community to be managed. Openly accessible data is often argued to provide the best ways for citizens to organize themselves around relevant issues and hold accountable those in power. Our research into one community’s gathering of data about flooding not only helped them to organize around the issue but also helped them to solve a recalcitrant

R. Beckwith (✉) · J. Sherry

Intel Labs, Jones Farm, 2111 NE 25th Avenue, Hillsboro, OR 97124, USA

e-mail: richard.beckwith@intel.com

J. Sherry

e-mail: john.sherry@intel.com

D. Prendergast

Department of Anthropology, Maynooth University, Room 2.15, Rowan House, Maynooth, County Kildare, Ireland

e-mail: david.prendergast@mu.ie

© The Author(s) 2019

M. de Lange and M. de Waal (eds.), *The Hackable City*,

https://doi.org/10.1007/978-981-13-2694-3_11

problem. However, we also found that making this data available as open data would lead to community impacts that were most unwelcome.

We will argue that in data governance for smart cities, the notion of ‘data as commons’ is crucial because community data is best understood as a rivalrous good that requires stewardship *by the community*. In addition, the notions of ‘datashed’ and ‘contextual integrity’ are presented as helpful in coming to a more nuanced strategy for the management of data and understanding of the affordances provided by data for communities. Simply put, we will argue that citizens of a smart city can find value in collecting and sharing data, but that they may also find value in restricting that data’s flow. Sharing and sheltering strategies will define data governance policies, which will, in turn, define how people can use that data for ‘hacking the city.’ We’ll close the paper with an argument that communities themselves must act as stewards of the data about their community and that sometimes this means that the data will not be fully open.

1.1 *The Value of Data*

The past decade has seen an explosion in the creation of—and interest in—data. Data had been growing in decades past, driven by individuals using the Internet and then mobile technologies. Most recently, we’ve seen volumes of data collected by digitally instrumented and connected devices. This superabundance of data has been called ‘The New Oil.’¹ This metaphor brings connotations of boomtown economics based on data flowing from a source to a purchasing destination. Indeed, most of the discussions of such data emphasize the financial returns and the importance of data acquisition. As one CTO has put it: ‘Even if I don’t know yet how I’ll use that data, I want it because I can store it so cheaply. My data science team might find a use for it.’ (Bertolucci 2014). The economics of data appear to be driving an explosion in surveillance undertaken by those large organizations with the reach and wherewithal to gather the most data. From this point of view, one could imagine a ‘smart city’ as a locus for the creation of new financial value for some favored few of its constituents. Given this, the city can be seen as a site of increasing surveillance—although, often, for no reason other than to enable a private entity to collect additional data for itself as it provides municipal services.

In contrast to this private acquisition-focused approach to ‘The New Oil’ is the Open Data philosophy, where data has no private owner and is made available to any and all. McKinsey Global Institute (Manyinka et al. 2013) argues that opening data up to broader sharing and use could generate \$3–5 trillion in economic value over the coming decade. Research suggests that these open approaches to data offer a variety of benefits. For instance, our own research in the Chilean *comuna* of Peñalolén

¹The quote ‘Data is the new oil’ has most commonly been attributed to marketing professional Clive Humby in a presentation at the ANA Senior Marketer’s Summit at the Kellogg School of Management, 2006.

showed that opening up city government procurement systems led to greater local participation in contracts, with more equitably distributed economic benefits (Kitner et al. 2007). In an entirely different arena, farmers using shared water data were able to demonstrate their ability to manage a watershed and avoid unwanted government intervention (Levin and Beckwith 2015). In a different arena yet again, Mann et al. (2002) have argued that ‘sousveillance’ by the populace, eyes on the powerful, would produce greater government accountability. Thus, freely shared (or open) data can have many positive effects.

For all of open data’s potential benefits, it is itself also a problematic construct, requiring us to ask such questions as who benefits and who might be harmed by the unselective sharing of data. Raman and Benjamin (2011), for instance, document what happened when Bangalore, India, put property ownership data online in the hopes of providing greater transparency and efficiency in property records. This inadvertently created a situation where those with the technical means and education were able to identify and effectively seize property that had problematic records. This enabled wealthier, more educated citizens to effectively steal land from citizens with less education, less technology access, or more tenuous legal claims on the property. Similarly, this chapter will address a smart city application focused on urban flooding. From one perspective, the open sharing of such data helped residents identify the source of the problem and organize for collective action. From another perspective, this community discovered that open flood data could, perhaps undeservedly, put some homeowners at risk of seeing property values suddenly and steeply decline.

In addition to these issues about who should share data and the potential impact on monetary value, there is also non-monetary value associated with data. Common models for dealing with smart city data do not seem to appreciate possible non-monetary values of data for the community (e.g., social value). This lack of awareness creates a sword that cuts two ways. On the one hand, the acquisitive private ownership model seems to see data only as material for ephemeral monetary transactions that have no history or future (Gudeman 2001). To this way of thinking, there are no relationships among people with which to be concerned. The community, to whom the data may refer, will have been forgotten. On the other hand, proponents of open data, in their rush to shed light on every aspect of a community, forget that communities consist of relationships and have boundaries. These relationships and boundaries help the community to cohere but are also vulnerable to forces from outside the community. We will see that sharing data can be detrimental to those relationships. Because of these issues, smart cities need more nuanced ways to think about data.

1.2 Thinking About the Flow of Data

Much of the recent interest in data is due to the fact that data has monetary value, but the value under discussion will accrue only if data flows. As we’ve noted, data can have both positive and negative values as it flows from one constituency to another. Given that data flow can create new value and can increase or decrease existing

values, we must ask: What data governance policies will best serve the citizens of a smart city?

Data doesn't flow by itself. It is pushed and pulled between different constituencies with their own goals and desires. Policies for data access and use create affordances that allow for these changes in value. Facilitations and constraints are placed on data flows, and these can determine the ways in which people can hack the city. Since it is the city's policies that create these affordances, they also must ask: How will these policies make our future cities 'hackable' in ways that citizens and communities desire?

In our thinking about how to construct a data governance policy for the people, we build on three conceptual frameworks: *the commons*, *datasheds*, and *contextual integrity*. These each inform our thinking about how smart city data *should* flow. The commons are community resources meant to be freely used by those in the community (in this discussion, that resource will be data). 'Datashed' is our term for all of the constituencies among whom some collection of data flows. Finally, 'contextual integrity' is a privacy framework (Nissenbaum 2004) that argues, in part, that people's expectations of information flow and use within a given context will determine their perception of privacy violations. Citizens' perceptions of privacy requirements for community data can be used to establish better policies (and regulations) for who should be able to use the data and for what.

1.2.1 The Commons

The commons is a well-known concept having to do with resources that are shared by members of a community: 'common pool resources.' Work regarding the commons (e.g., Ostrom 1990) is important to consider, especially because recent years have seen a very reasonable push to make civic data 'open.'²

Open data has often been said to establish a 'data commons' (e.g., Grossman et al. 2016). Commons resources are considered public goods, meaning that they are accessible to the public, and also *rivalrous*, meaning that their use by one precludes their use by another. Rivalrous phenomena are contentious because of the potential diminution of the value of the resource for later users. As an example, the grass in a shared grazing land: If one person's cattle eat all the forage, there will be none left for the cattle of others. Because of the rivalrous nature of common pool resources, they need to be protected from overuse. A key focus of Ostrom's studies of the commons is how non-market mechanisms are used by communities (and not a remote government or local gentry) to enact stewardship and to ensure sustainability of such

²Two examples: (1) By Executive Order, the US government ("Making Open and Machine Readable the New Default", 2013) has mandated that 'Government information shall be managed as an asset throughout its life cycle to promote interoperability and openness, and, wherever possible and legally permissible, to ensure that data are released to the public in ways that make the data easy to find, accessible, and usable.' (2) Open data is described by UK-based Open Knowledge International: 'Open data and content can be **freely used, modified, and shared** by **anyone** for any purpose' (emphasis in original).

resources (1990). Research into stewardship has established the deep intermingling of resource management and the community's social and cultural practices (Netting 1981; McKay and Acheson 1990).

One of the conceptual challenges of considering open data as a commons issue arises from the fact that the notion of data 'ownership' is fraught.³ Consider that data is often created at points of interaction among multiple parties—at the point of purchase, for instance, involving a buyer, a vendor, and a credit card company, all of whom may feel some entitlement to transaction data. All three are actors in the sales event. Clearly each of these actors has the potential to claim ownership. Data, therefore, often has ownership claims distributed across a number of parties. Dealing with these claims is one of the roles of a smart city.

Data *about* the commons increases this challenge. Should a private party be able to exclude community members from seeing data that the private party has collected about a community resource? For a negative example, consider whether a London cabbie (or London Taxi and Private Hire, which oversees the test for 'The Knowledge' of the arcane London street map) should be allowed to stop people from using GPS-enabled mobiles with maps because cabbies have traditionally been associated with The Knowledge. This kind of restriction is certainly not in the service of greater London (or anyone aside from cabbies) and wouldn't be likely to find much support, legal or otherwise. Maps of public thoroughfares can be owned but not the right to map. We might ask whether a private party could withhold from public view any data about 'public' resources. Consider, for example, privately collected data related to a grazing ground or even weather data. Should private companies be allowed to collect such data and keep it private? Examples from our fieldwork (reviewed below) suggest that the answer is not so simple.

In addition, questions about the rivalrous nature of information resources themselves raise another difficulty. In some ways, it is compelling to find a conceptual difficulty considering open data as a commons issue. Digital data can be copied endlessly with no diminution to the original in physical terms. Unlike most material goods, data and information are often considered non-rival goods—their access or use by one party does not preclude access or use by others (Benkler 2004).⁴ We believe that, in rivalry, the value of the resource is key. While it can be argued that data copies easily without changing the ability to physically access that same data for another user, access and monetization of the data do not exhaust the values that a piece of information might have. In fact, information (the stuff of open data) has been argued by Aragon (2011, discussed below) to have at least three forms of value—economic, sociological, and identity. The diminution of *any* of these values due to circulation, then, demonstrates that information resources are potentially rivalrous. Stewardship of the data itself, to which we shall now turn, is how communities can preserve those values.

³Indeed, Bezaitis and Anderson (2011) argue that, in the context of so many new information technologies, the very concept of ownership is in a state of flux.

⁴See Benkler's (2004) exegesis of non-market production of digital information and the results of the placement of that information into the commons.

Stewardship. Elinor Ostrom was awarded the Nobel Prize for her work on stewardship and the commons. Before her work, many economists had been swayed by potential overuse—‘the tragedy of the commons’—and argued for the rationality of removing rivalrous shared resources from the common pool (e.g., through private ownership of these resources) (Hardin 1968). Instead of privatization, Ostrom showed through her work investigating communities where the commons were left to the community that local stewardship could be effective. Ostrom offered eight ‘design principles’ (1990) that were present when communities could effectively engage in stewardship. These design principles were: (1) well-defined boundaries, (2) broad compliance with shared stewardship practices within those boundaries, (3) locally relevant stewardship rules, (4) effective compliance monitoring, (5) appropriate sanctions for non-compliance, (6) mechanisms for easy arbitration, (7) broad recognition of local powers, and (8) tiered management for large resources. When most (but not necessarily all) of these are in place, a commons can be effectively and sustainably managed from within.

One well-known example of effective commons stewardship involves the lobstermen of Maine as they worked together to manage fishing practices to ensure a sustainable lobster population (Acheson 2003). This example embodies many of the principles Ostrom noted as necessary to protect the commons. Here, the threat of overfishing lobster in local estuaries could impact livelihoods so government regulation was proposed as a reasonable strategy for ensuring that lobsters would be plentiful. The community resisted outside regulation. To forestall regulation, the fishing community drove a set of relationships and agreements among various constituencies, including dealers, legislators, conservation groups, and state agencies (among others) to develop a set of institutional practices. These practices were developed to protect a common pool resource, that is, the lobsters around the Maine coastline.

Ostrom’s principles were well represented here. In this case, (a) lobstermen and parties with economic and ecological interests in their activities, (b) within a specific state of the USA and a region within that state, (c) saw the threat of a reduction in the output of lobster, and (d) they developed rules that could be easily enforced through sales channels. This locally driven approach proved to be remarkably effective.

Stewardship of information. Research on the commons has also been applied specifically to the use and sharing of information (Kollock and Smith 1999). When we consider how stewardship of that information should be accomplished, we must look to the community itself for local guidance because the ways in which a particular community may choose to enact stewardship can vary in surprising ways. Aragon’s (2011) work provides a specific example of the ways in which different communities steward similar information differently. Aragon looked at the surprisingly dissimilar ways in which two communities control the flow of similar information in order to steward their respective cultures. She frames her discussion in terms laid out by Gudeman: ‘taking away the commons destroys community, and destroying a complex of relationships demolishes a commons’ (2001, 27). That is, the commons and the specific community that shares it are inseparable. Aragon argued that controlling the flow of information is one way that communities express and steward their culture. Considering how communities choose to steward their culture (and their shared

information) allows us to see that it is not just the information but also shared beliefs about that information that define the practices of data governance.

Aragon compared two communities that manufacture textile goods and the different ways that they handled information about how these goods were produced. One employed a ‘circulation’ strategy in which they were happy to have outsiders gain access to the knowledge of the methods that they use to produce the goods. Another employed a ‘sequestration’ strategy where they tried to keep production methods a secret outside of their group. Their choice of strategy depended upon what type of value people were trying to steward. In the first case, the community felt that if their knowledge was kept alive, that would keep their culture (and community) alive so they chose circulation. The second community feared that if outsiders shared the knowledge of how they produce their goods, then the outsiders could steal their relationships with customers and their community would be diminished, so they chose sequestration. These contrasting strategies for stewardship—circulation and sequestration—are valuable concepts to use when we think about how a community wants to share data. It is worth noting that what is called ‘circulation’ here is the typical notion of open data. Sequestration, though, does allow for some data flow, but the flow is limited only to those inside a defined community.⁵

1.2.2 Datasheds

As described by Ostrom, a key element of successful management of a commons is a clear sense of physical boundaries. In talking about data circulation, one must address the boundaries within which data circulates. This is what Levin and Beckwith (2015) called a ‘datashed.’ Just as a watershed helps hydrologists think about water, looking at the circulation of data—its datashed—helps us to think about civic data. Because information shifts in value as it flows, observing the sites to which data flows tells us about how value may be assigned, who collects the data, and also tells us more about what those people care about.

Levin and Beckwith (2015) examined a community where a recent initiative had sought to use ‘Internet of Things’ (IoT) technologies to instrument a wide variety of industries and sectors. The data generated by these technologies was meant to flow not only between various constituencies with a history of interaction (e.g., among local farmers and the truckers who move their crops) but also to bring in new players who may have an interest in the data (e.g., investors in commodity futures or the banks that loan money to farms). That is, data would not just be used by the collectors and those with whom they collaborate to bring a product to market; the data would also be used by people within the same or adjacent industries and even people interested

⁵Not collecting data at all is a strategy, too. Some Native American communities do not collect or map the sacred sites for tribal members and, as a consequence, the tribes cannot share such information with those who would seek to develop the lands. What’s important to note here is that communities make decisions about data flow. Communities act (either as a collection of individuals or in concert) as the owners of the data.

in the data for purposes entirely distinct from the original intent. The datashed would include all of these people.

Levin and Beckwith called the value of data as it circulated outside the initial site or original intent ‘circulatory value.’⁶ Circulatory value has implications both for ‘sheltering’ and ‘sharing’ approaches. Positive circulatory value (for sharing) will depend upon the existence of an alternative constituency which may or may not have a common interest. Data only have value when their use or restrictions on use help someone achieve a goal. Once we understand this, it becomes easier to see why people often have concerns about downstream recipients of data, especially when that other’s goals are incommensurate with their own. This is where expectations of privacy come in and why we think it’s important to consider contextual integrity.

1.2.3 Contextual Integrity

Contextual integrity is the privacy framework that we used to think about the role of communities in data governance decisions. Contextual integrity (or Privacy in Context) (Nissenbaum 2004) provides a structure for addressing issues around stewardship by allowing people’s expectations of privacy to shape the rules for information flow. Contextual integrity establishes a framework for the problematic challenge of ensuring privacy in a society where new information technologies enable an ever-increasing sphere of public surveillance. Contextual integrity uses a concept quite like datasheds called ‘contextual boundaries.’⁷ Individuals define these contextual boundaries to contain the entities to which they believe their personal information might reasonably flow. The boundaries exclude entities to which the data should not flow. Through contextual integrity, we are able to identify a number of lenses through which to consider the ‘sharing’ or ‘sheltering’ of civic data. Within the framework of contextual integrity, Nissenbaum talks about three roles that people might fill with respect to shared personal information: information receiver (the person to whom data is transferred), the information sender (the agent acting to transfer the data, to cause it to flow), and the subject (the entity whom the data is ‘about’).

Nissenbaum’s work has been primarily applied to issues of personal data and privacy, but it is also a useful framework for thinking about the circulation of civic data. Specifically, combining the concept of contextual integrity with an understanding of civic data as a common pool resource, we can ask how community members, in addition to municipal governments or other large institutions, might contribute to and interact with data and information that is deemed valuable by the community. What facilitations and restrictions on gathering and use need to be applied? How should flow be controlled among community, municipality, and state? What about private

⁶This circulatory value, when considered in the context of Aragon’s work, could be the value of having one’s culture survive.

⁷Datasheds focus on the places where data flow. There is no sense in which the goals or values of constituencies are reflected. Contextual boundaries, however, do address goals and reflect desires with respect to data flow.

enterprise? It also raises questions such as what are the settings in which data might be appropriately gathered, who might legitimately lay claim to such data, and under what circumstances might it be circulated?

Before getting to our case study, we should review the three areas we believe are important for thinking about them. We have reviewed work on the commons showing how local governance can lead to sustainable resources. We talked about data flow and how various constituencies may interact with a set of data within what we are calling a datashed. Finally, we explored how rules for flow might be constructed so as to preserve contextual integrity—privacy.

We will now turn to a focused case study to help us understand data governance for a smart city. In addition to highlighting the importance of situatedness, the example below demonstrates the ways in which data or information can bring together opposing constituencies. In this particular case, it happened that some of those brought together by the data were somewhat unwelcome by others. In addition, and as a consequence of those unwelcome others, this example also provides a clear example of where a community wants to withhold data about the commons from others. It is our contention that the problems occurred because the interpretations of the data by remote users of the data were at odds with the understanding of the data shared by local community members, whose situated knowledge provided a different understanding.

2 Case Study: Watersheds and Datasheds

This case study concerns a US suburban town that had recently developed a significant problem with flooding. We worked with residents over a two-year period where we also spent time with government agencies that were undertaking activities in the community. We also worked closely with an advocacy group that was trying to influence policy and funding in the community.

We spent considerable time with one woman, in particular, who had lived in her house for over 25 years. In more recent years, her home had flooded over ten times. She was initially told by local government officials that there was no change in flooding within the community and this was a problem that was hers alone. Based on the fact that she lived hundreds of meters from the stream that was flooding and that a lake regularly formed in the backyards of all the people on her block, she knew this was not her problem alone. She described for us how she set about trying to get her neighbors involved in finding a solution. She canvassed the neighborhood and found others, like her, who were suffering property damage from an increasing number of floods. She enlisted these others to help the community understand more about the new floods. The group decided to create a map of each flooding event. With their mapped data, they were able to demonstrate that there was a significant flooding problem across their community and again asked the local government for help.

Even after collecting the data and sharing it with town officials, she and her neighbors were told that there was nothing that this group or even the town could

do. The officials claimed that this flooding was caused by climate change. It was, in effect, the new normal. This narrative held that because of changes in patterns of precipitation, the existing infrastructure was no longer capable of supporting the runoff and that changes in infrastructure would need to be balanced against other municipal expenditures. The community group did not believe this explanation and felt that infrastructural changes in an upstream community were to blame. These infrastructure changes were well known to this group, and they had a theory of exactly how it might have influenced flooding in their community. Their theory was supported by the data that the group collected.

The potentially relevant upstream infrastructure changes were discovered because another aspect of the group's work was to try to find the water sources, so they explored the full upstream watershed during flooding events. They found the locations where the stream flow began to increase substantially. There was one spot at a golf course where, they discovered, the culvert leading from the course was recently cleared of brush to facilitate drainage into the head of the stream. Another spot was the site of recreational sports fields that had been built in the past few years. A retention pond was built to compensate for the change in water flow that was caused by the sports fields, but this group observed that the pond was not filling during flooding events. These facilities were not in the same town but in an adjacent town where the flooding stream originates, and the incomes and property values are higher.

The group also tried to see what kinds of government programs were available to their community and to share this information with their flood-mates. This was when they discovered that certain federal money would be hard to get. It seems that, according to the Federal Emergency Management Authority (FEMA) maps, they were not in a floodplain. This did not preclude getting government funds but made these funds more challenging to access.

Their town had no jurisdiction in any case since the problem originated upstream. This exemplifies an interesting property of datasheds: A datashed is not necessarily coextensive with a single jurisdiction. A community can choose to extend its datashed well beyond the community's jurisdictional boundaries. This group pushed beyond the officials of their town and sought relief from regional and national agencies charged with stewardship of the waters. When they tried to see if there was something that could be done to protect their downstream community, they were informed that they were ignorant of the situation and lacked credentials required of someone who could understand a watershed. Nevertheless, they had a body of theory, data, and maps which they subsequently brought to many public meetings.

At one of the more raucous public meetings, there was a representative of the upstream community that the residents blamed. He was quiet through much of the meeting but when residents started to complain about his town and blame the upstream community for the floods, he stood up and informed the group that he worked for the town and was, in fact, the person in control of the retention pond and that it, too, was overflowing during flooding events. The group then produced photographs they said were taken during floods that showed the pond was not filled as it should have been. He questioned whether the photographs were actually taken when the residents claimed. This photographic evidence was open to being questioned,

but the accusation was now out there. Interestingly, after this meeting, whenever there was a heavy rain, the group would go and check the retention pond, and it was always full. More interestingly, the flooding also abated. It would seem that the residents were right. Despite their lack of hydrology credentials, they were able to use their awareness of local conditions to collect relevant data and interpret this data in a manner unavailable to their credentialed but remote partners.

The story is not yet over. This community next faced a new problem. Recall that FEMA maps did not have this community as a floodplain. Across the USA, FEMA is in the process of redrawing the flood maps that it uses to assign risk to communities. Existing maps are inaccurate and insufficient, but it is expensive to collect new data. To what extent should the data that was collected to argue for these successful mitigation strategies be used to characterize the flooding potential of the community? Recall that flooding in this community was felt to be a function of upstream mismanagement, a problem that has been rectified. The homes are not flooding as they were. What FEMA would like to do is to use the data collected by this community to determine the level of risk to assign. Obviously, if they use that data without considering that potential causal factors have been addressed, they will determine that a large number of people need to carry flood insurance. This insurance could add about 20% to the average monthly mortgage payment and potentially reduce the value of homes. Community members feel that this is unfair as the data had been used to fix the problem, and they decided that they were no longer willing to share data with the federal government. That is, they developed a sequestration strategy.

2.1 Circulation and Sequestration

While free circulation—that is, open data—is a popular option for data from the smart city, sometimes data may be better suited for a ‘commons-like’ treatment. A more suitable option may be free use within the community, but sequestration of that data with respect to some parties or for some uses outside the community. With this in mind, we address sequestration with respect to data about the commons.

We might ask first, what are the boundaries of the commons? The datashed, watershed, and jurisdictional boundaries can all be dissimilar. That is, the boundaries of one may not be the boundaries of another. The first data flow option to occur to a community might be to allow data to circulate freely to enable openness and accountability. However, expectations around data flows are important to understand. Contextual integrity tells us that we should be especially concerned with the expectations of those whom the data is about. We believe that the ‘subjects’ of commons data are community residents, those locals charged with stewardship of the physical resources of the commons. This militates against the notion that all potential constituencies of the datashed should have equivalent access to the data or equivalent power in determining data flows.

One point to consider here is that the datashed is sometimes not the same as the resource boundaries because the resource may be controlled by actors who are outside that boundary. Frequently, elements of jurisdiction or control over a resource are a function of distant parties, and in these cases, data sometimes must be shared with these distant participants. The datashed, then, cannot be constrained to the entities within the boundaries of the resource. When distant authorities regulate local resources, they may use locally collected data as a tool. What we show here is a case where the locals who collect the data want to sequester the data from some distant authorities who are desirous of regulation.

As noted, open data circulation can be quite beneficial. However, it is also the case that sometimes people do not want specific data to circulate freely or to share that data with specific others. For example, misleading data that is consistent with frequent flooding or even the risk of flooding can be used to mandate that home owners carry significant flood insurance which can impact the value of a home. It might come as no surprise that some people are hesitant to share information. They don't want open data—maybe just *slightly ajar data*. Some people might argue that anything less than full disclosure of this information is dishonest. What if the data being shared would easily invite inferences that are incorrect?

The costs associated with sharing are a consideration for people in the community. Even before the time that the upstream problem had been addressed, let alone FEMA threatening to reduce the value of their homes, one community member told us that some 'people are always afraid that it's going to be "information means punishment".' It is not that they do not wish the problem solved, they are simply afraid that they will ultimately not benefit from data sharing.

Sequestration does not mean that there can be no sharing at all. These people were happy to share their data with those involved in mitigation. The sequestration that they argued for would restrict the parties among whom the data would circulate and the purposes to which the data could be put. This request is not out of line with how we would expect stewardship to come into play around data that a community has willingly collected. It hardly needs to be said that an unwillingness to participate in sharing is quite problematic from the perspective of open data. If people do not participate, there will be no data to make open.

2.1.1 Outstanding Problems

Community-led circulation and sequestration decisions may not work to support every individual. Consider a person who wants to opt out of the flood information system because they do not ever want to share their flooding status with anyone. What if they are right in the middle of flood zone? A system using local topography and the presence of water in some locations could clearly implicate their property as one that would be inundated before a neighbor's (higher) property. How can such a person opt out? It really is not an option. Inferences can be drawn from a neighbor's data. With open civic data, there may be no way to truly implement an individual 'right to be forgotten' since the inferences across the commons are made irrespective of

the individuals. Yet, properties associated with individuals may be easily identified. Figuring out governance issues such as how to protect or whether to protect data in such a system will be important.

One of the issues with civic data is that, by providing transparency, this data can support accountability. That was certainly the case in Peñalolén where community residents were finally able to profit from municipal procurement (Kitner et al. 2007). It was easier to see when money was being spent and whether favoritism was involved in vendor decisions. Accountability, in fact, is often held up as one of the most important outcomes following from open data. However, one person's accountability can be another's control. By making visible the results of one's actions, this could invite inferences about activities or states that one might prefer not to imply. Sensors cannot show that reasonable decisions have been made for reasons outside the view of the sensors. If interpretation of data requires contextualization that is not available to all data users, how is that accountability?

Another issue with open city data is something that we have seen widely throughout the IoT developer world. Many denizens of the datashed are not capable of managing the data science to produce answers to the questions they would ask. Others may be vulnerable to exploitation by tech elites as we saw earlier in the Bangalore example (Raman and Benjamin 2011). This lack of data science expertise means that some people will not know how to meet their needs relative to the circulating data. This does not mean that they will not be part of the datashed. In fact, people may not have an option; the data may implicate them in any case. What this lack of data literacy means is that some people will need to enroll others in the datashed who will educate, represent, or collaborate with them.

3 Discussion and Summary

As we think about hacking a smart city, it is wise to think about what a smart city does. At smart cities' core is the creation and use of data for new services. Many proponents of smart cities encourage the idea that this data should be made open to support a new economy. The main argument of this chapter is that smart cities have a choice of what to do with their data; information resources can be open and available to all or they can be understood and managed as a commons. There are significant differences between these two options. On the one hand, open data is typically free to all with no owner controlling the flow of data. On the other hand, a data commons, as is true for all commons, should be about resources held in common by a group. A data commons effectively asserts group ownership of the information resources. This data would, of course, be collected and distributed to benefit that group.

Rivalry, Stewardship, and the Commons. We argued that the shifts in value that follow from data flow allow us to conceive of information as rivalrous and, thereby, characteristic of what stewardship of the commons is meant to manage. The changes in value we've referred to have to do with value being created or destroyed as data flow from one constituency to another. If value for the first constituency can

be lessened by the flow of data to a second constituency, then we have a form of rivalry.

Because of this potential rivalrous nature of data, stewardship of informational resources will be key to a successful data commons. A commons has the property that the community has to take care of (or steward) the resource. The community has to worry about sustainability and equity and must, in the case of information resources, put data governance procedures in place that will ensure these.

This means that cities have a social role to play with stewardship of data. Data governance needs to consider data gathering, analysis, sharing, sheltering, and is, above all, necessary so that the community has control over shifts in value. They will be concerned with the circulation and sequestration of data flows both inside the city and when data leave the city. Stewardship is not just about the data while it is in the city. Stewardship is for the life of the data throughout the datashed.

In addition, we argued that while smart city data can create community, not all communities are created equally. When needs are met and value created by broadly accessible affordances, various constituencies can be formed, some with ties to the community, some without. Within the city, constituencies will mostly be composed of those with some kind of relationship to the resources or phenomena being measured. Although the data is produced by measuring phenomena within a community, other constituencies can be brought into the datashed by data alone. Because their interest is only in data, these outside communities can more easily have interests that conflict with those stewarding the measured resource.

We have argued that stewardship is one of the social roles that the smart city must play. There is another way to look at data stewardship that is perhaps even more obviously social. Stewards of information resources must understand the range of players and consider the pathways and consequences of how data will be used. A smart city is the most likely candidate to define the criteria for admission to the datashed and communities need to trust city administrators to put good data governance in place. In making decisions about sharing, they may be 'inviting' outsiders to interact with city resources.

Who Owns the Data of the Smart City? In terms of real ownership, it might be reasonable to say that no one owns a data commons. That is why no individual is empowered to sell to another that thing over which they are the steward. The most important aspect of a steward is that their job is to take care of a resource so that its value does not diminish for others.

Given that the same data could have both positive and negative impacts, it is important to ask who should make the stewardship decisions and whose positive and negative impacts should have priority. Not everyone who might have a stake *in the data* should have an equivalent say in data flows. Obviously, people who see value for themselves in the data would be interested in how the data flows but an interest in extracting value does not give them the right to control the data. Also, among those with a stake in the data are those who see a potential diminution of value (of the measured phenomena, the data, or the community at large) when that data moves beyond a 'contextual boundary.'

Among each of the potential constituencies, just one can make the final decision as to which values must be preserved and which flows must be forbidden. Whose values are most significant? We know from the work in contextual integrity that when data is *about someone*, expectations of privacy are most significant. Perhaps, then, the question should be ‘who are the data about’? In many ways, the data could only be ‘about’ a community that knows how the data relates to the measured phenomena, people who know how to interpret the data as it relates to the local resource.

Data is ‘about’ locals, since they are best able to understand the data and its meaning. Some potentially impactful interpretations of data actually require situated knowledge, the requirement for which impairs the distant communities’ interpretation of local data. In our fieldwork, for example, the implications of circulating data (without situated knowledge) could be seen as negative and unfair.

We have tried to show that typical ‘smart city’ data—data about the commons—may require restrictions on data flow. As we’ve seen, openness of data may not always be the best thing for a community nor what a community might choose for itself. Circulation and sequestration are data stewardship strategies that need to be considered with smart city data. Whatever strategy is chosen, processes need to be put in place for decision-making that are consonant with community desires. Then the stewardship of information resources can help people to work together. This is one way that communities can cohere.

Smart cities can be a locus for the creation of new value for those within the city. They can also be the locus of serious breaches of trust where information can be shared to provide value to others while it simultaneously harms city residents. As a bulwark against this, we believe that a city should manage its data as a commons. To do so means trying to understand potential data flows and the values of the communities within the city, while also being respectful to rightful claims of ‘ownership’ and rules of stewardship. If cities do this, they can expect that the citizens of the smart city will be better served by the smart city itself and will be more strongly invested in its success.

References

- Acheson, James M. 2003. *Capturing the commons: Devising institutions to manage the Maine Lobster Industry*. Hanover: University Press of New England.
- Aragon, Lorraine V. 2011. Where commons meet commerce: Circulation and sequestration strategies in Indonesian arts economies. *Anthropology of Work Review* 32 (2): 63–76.
- Benkler, Yochai. 2004. *The wealth of networks*. Online Open Publication. http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf. Accessed 18 Apr 2017.
- Bertolucci, Jeff. 2014. *When data hoarding makes sense*. *Information network*. <http://www.informationweek.com/big-data/big-data-analytics/when-data-hoarding-makes-sense/d/d-id/1297474>. Accessed 18 Apr 2017.
- Bezaitis, Maria, and Ken Anderson. 2011. Flux: Creating the conditions for change. In *Proceedings of ethnographic practice in industry conference proceedings 2011*, 12–17.

- Grossman, Robert L., Allison Heath, Mark Murphy, Maria Patterson, and Walt Wells. 2016. A case for data commons: Toward data science as a service. *IEEE Computing in Science and Engineering* 18 (5).
- Gudeman, Stephen. 2001. *The anthropology of economy: Community, market, and culture*. London: Blackwell.
- Hardin, Garrett. 1968. The tragedy of the commons. *Science* 162 (3859): 1243–1248.
- Kitner, K., R. Beckwith, and N. Boaitay. 2007. Optimizing cultural and economic security in the implementation of digital development: The case of Penalolen, Chile. In *Proceedings of the 9th international conference on social implications of computers in developing countries*. São Paulo.
- Kollock, Peter, and Marc A. Smith (eds.). 1999. *Communities in cyberspace*. New York: Routledge.
- Levin, P., and R. Beckwith. 2015. Datasheds and the remaking of agricultural practices: How the internet of things is changing agriculture. In *Society for economic anthropology conference*. Lexington.
- Making Open and Machine Readable the New Default for Government Information. 2013. www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government. Accessed 26 July 2018.
- Mann, Steve, Jason Nolan, and Barry Wellman. 2002. Sousveillance: Inventing and using wearable computing devices for data collection in surveillance environments. *Surveillance & Society* 1 (3): 331–355.
- Manyinka, James E., Mihale Chui, Dianna Farrell, Steve Van Kuiken, Peter Groves, and Elizabeth Almasi Doshi. 2013. *Open data: Unlocking innovation and performance with liquid information*. McKinsey Global Institute. <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. Accessed 26 July 2018.
- McKay, Bonni, and James Acheson (eds.). 1990. *The question of the commons: The culture and ecology of communal resources*. Tucson: University of Arizona Press.
- Netting, Robert McC. 1981. *Balancing on an Alp: Ecological change and continuity in a Swiss mountain community*. Cambridge: Cambridge University Press.
- Nissenbaum, Helen. 2004. Privacy as contextual integrity. *Washington Law Review* 7 (1): 119–157.
- Ostrom, Elinor. 1990. *Governing the commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press.
- Raman, Bhuvanawari, and Solomon Benjamin. 2011. Illegible claims, legal titles, and the worlding of Bangalore. *Revue Tiers Monde* 206: 37–54.

Richard Beckwith is a Research Psychologist in Intel Labs with the Adaptive Computing Lab's User Experience Insights. Richard has been with Intel since 1996. He has published primarily on language, education, sensors, and privacy. His recent work has focused on service development for community-based sensing schemes. He received his Ph.D. from Teachers College, Columbia University in 1986. His degree is in Developmental Psychology and his work focused on how statistical parsers could enable language acquisition. From 1986 to 1991, he was a research scientist at Princeton University's Cognitive Science Lab, working on WordNet. Before coming to Intel, he was an Associate Professor at Northwestern University in the Institute for the Learning Sciences (ILS) where, from 1991, he taught qualitative methods to support technology design and oversaw the development of assessment schemes for AI-based educational software.

John Sherry is a social anthropologist and manager of User Experience Insights at Intel Labs. John joined Intel in 1997 as the company's first anthropologist, serving as a founding member of what was then called Intel's People and Practices Research Lab. Subsequently he has served in a variety of positions, including director of User Experience Design in Intel's Digital Health Group, before taking on his current role. His research has always involved understanding ordinary people in their everyday settings, including the complex social, economic and technical systems in

which we are all embedded. The purpose of this research is to help Intel better imagine and invent new uses and experiences of computing. He holds a Ph.D. in Anthropology (1995, University of Arizona) and a B.S. in Computer Science.

David Prendergast is a social anthropologist and Professor in Science, Technology & Society at Maynooth University in Ireland. His research career over the last twenty years at Cambridge, Sheffield, Trinity College Dublin and Intel has resulted in a wide range of books and articles on ageing, health, cities, science, technology and social relationships. His most recent volume 'Aging and the Digital Life Course' edited with Chiara Garattini was given a CHOICE 'Outstanding Academic Title' by the American Library Association and was described as one of 'the two most fascinating books on aging in the 21st Century' by the Huffington Post.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

