7

Shaping participatory public data infrastructure in the smart city

Open data standards and the turn to transparency

Tim Davies

Introduction

One of the earliest expressions of the modern open data movement evolved out of the CitiStat analytical dashboard in Washington DC. In 2007, seeking to allow actors outside the administration work with city data, the DC chief technology officer launched data.dc.gov, a data portal offering direct access to machine-readable government datasets (Tauberer, 2014, Ch. 1). Over the following decade, the idea that governments should publish their data holdings as machine-readable, freely accessible and openly licensed data has taken hold across the world at national and sub-national levels (Davies *et al.*, 2019). Open data ideas have found particular traction in urban areas, initially connecting with a culture and practice of civic hacking (Landry, 2019), and open data has also acted as a component of a number of smart cities programmes, both rhetorically and substantively. Scholars have been particularly interested in the potential uses of open data to support democratic engagement and collaborative models of governance in the smart city (Bartenberger & Grubmüller-régent, 2014; Goldsmith & Crawford, 2014). Yet open data communities have also been the source of a number of critical perspectives on the smart city: questioning centralisation and corporate control of urban data infrastructure, and challenging the presentation of a rationalised urban domain filled with consumers and service recipients, rather than a rich urban environment of diverse citizens, political struggles and lives only partially digitised (e.g. Sadoway & Shekhar, 2014).

One source of this dual role of open data, in both enabling and opposing smart city narratives, can be traced to the complex origins of the open data movement, which brought together both public sector information businesses and civil society activists, united by a common cause in gaining access to government data, but ultimately motivated by divergent long-term goals (Gonzalez-Zapata & Heeks, 2015; Gray, 2014). These unusual allies ranged from those seeking transparency, accountability and new forms of civic engagement (Davies, 2010; Huber & Maier-rabler, 2012; Kassen, 2013; Sieber & Johnson, 2015), to those looking

to develop new business models and promote the idea of 'government as a platform' outsourcing many more aspects of service provision to the private sector (Gurin, 2014).

In this chapter, my goal is to explore open data-related strategies for re-asserting the role of citizens within the smart city. Such strategies are able draw in particular on the political narrative of transparency, and the role of technical standards in delivering transparency. I will outline how these two components can be used not only to secure access to data from government, but also to open up two-way communication channels between citizens, states and private providers. I will argue that a focus on opening up the data infrastructures of the smart city not only offers the opportunity to make processes of governance more visible and open to scrutiny, but it also creates a space for debate over the collection, management and use of data within governance. This can give citizens an opportunity to shape the data infrastructures that do so much to shape the operation of smart cities and of the modern data-driven policy environment.

The chapter proceeds in four parts, the first three unpacking different aspects of the title, and the forth offers a model for thinking about the relationship between transparency, open data and standards in the future development of inclusive and participatory data practice in the smart city.

Part 1: Participatory public data infrastructure

Data infrastructure

Infrastructures provide the shared set of physical and organisational arrangements upon which everyday life is built. The notion of infrastructure is central to conventional imaginations of the smart city. Fibre-optic cables, wireless access points, cameras, control systems and sensors embedded in just about anything, constitute the digital infrastructure that feed into new, more automated, organisational processes. These, in turn, direct the operation of existing physical infrastructures for transportation, the distribution of water and power, and the provision of city services. However, between the physical and the organisational lies another form of infrastructure: data and information infrastructure.

Although in the literature the term 'information infrastructure' is often used to cover both data and information, I use the two terms separately here to draw attention to an important analytical distinction. The General Definition of Information (GDI) describes information as 'data + meaning' (Floridi, 2004). Information, as the basis for human decision-making, requires data that is filtered, organised and contextualised. Data, by contrast, is, in its purest form, decontextualised: with each individual aspect of a phenomena encoded as a distinct data point, open to be re-assembled and represented as information, but also open to a range of different representations and forms of analysis. It is this re-interpretability that gives digital data its particular value. Of course, in practice a digital dataset rarely encodes all the possible variables that describe a phenomenon; instead, certain features of the world are selected for encoding and others discarded. Even with growing data storage and processing capacity, the need for this explicit or implicit selection is not avoided.

It is by being rendered as structured data that signals from the myriad sensors of the smart city, or the submissions by hundreds of citizens through reporting portals, are turned into management information and fed into human- or machine-based decision-making, and back into the actions of actuators (Dunleavy *et al.*, 2006) within the city. Seen as a set of physical or digital artefacts, the data infrastructure of a city involves ETL (Extract, Transform, Load) processes, APIs (Application Programming Interfaces), databases and data warehouses, stored

queries and dashboards, schema, codelists and standards. Seen as part of a wider 'data assemblage' (Kitchin & Lauriault, 2014) this data infrastructure also involves various processes of data entry and management (Denis & Goëta, 2017; Goeta & Davies, 2016), of design, analysis and use, as well relationships to other external datasets, systems and standards. Dodds and Wells capture this by defining data infrastructure to incorporate not only data assets, such as datasets, identifiers and registers, but also the organisations and organisational processes used to provide access to those assets (Dodds & Wells, 2019)

It is, however, often very hard to 'see' data infrastructure. By their very natures, infrastructures move into the background, often only 'visible upon breakdown' (Star & Ruhleder, 1996). For example, you may only really pay attention to the shape and structure of the road network when your planned route is blocked. It takes a process of 'infrastructural inversion' to bring information infrastructures into view (Bowker & Star, 2000), deliberately foregrounding what has been so far the background. I will argue shortly that 'transparency' as a policy performs much the same function as 'breakdown' in making the contours of infrastructure more visible. In taking something created with one set of use-cases in mind, and placing it in front of a range of alternative use-cases, transparency allows the affordances and limitations of a data infrastructure to be more fully scrutinised. Such critical scrutiny can then feed into shaping the future development of that infrastructure. But before developing that argument further, I will first outline the extent of 'public data infrastructure' and the different ways in which we might understand the idea of a 'participatory public data infrastructure'.

Public data infrastructure

A city may have a wealth of public data yet without a coherent public data infrastructure. In *The Responsive City*, Goldsmith and Crawford describe the status quo for many as 'The century-old framework of local government – centralized, compartmentalized bureaucracies that jealously guard information' (Goldsmith & Crawford, 2014). In such a city, datasets may exist, but they are disconnected. For example, one department may maintain detailed digital maps showing the land the city is responsible for maintaining and city assets, whilst another manages contracts with outsourced maintenance crews, who work from entirely separate maps. Even so, such a city may have come to publish some data online in an open data portal in response to transparency edicts, but each dataset exists as an island, published using different formats and structures, without any attention to interoperability

It is against this background that initiatives to construct public data infrastructure have sought to introduce shared technology, standards and practices that provide access to a more coherent collection of data generated by, and focusing on, the public tasks of government. For example, in 2012, Denmark launched their 'Basic Data' programme, looking to consolidate the management of geographic, address, property and business data across government, and to provide common approaches to data management, update and distribution (Government of Denmark, 2012). In the European Union, the INSPIRE Directive and programme has been driving creation of a shared 'Spatial Data Infrastructure' since 2007, providing reference frameworks, interoperability rules and data sharing processes (Bartha & Kocsis, 2011). And more recently, the UK government launched a 'Registers programme' (Miller & Roe, 2018) to create centralised reference lists and identifiers of everything from the names of countries to a list of government departments, framed as part of building governments digital infrastructure. At the city level, increasing adoption of smart city technology, particularly when cities choose to work with a range of vendors, is likely to drive greater development of urban data infrastructure and frameworks for city data interoperability.

The creation of these data infrastructures can clearly have significant benefits for both citizens and government. For example, instead of citizens having to share the same information with multiple services, often in subtly different ways, through a functioning data infrastructure governments can pick up and share information between services, and could provide a more joined up experience of interacting with the state. By sharing common codelists, registers and datasets, agencies can end the duplication of effort and increase their intelligence, drawing more effectively on the data that the state has collected.

However, at the same time, these data infrastructures tend to have a particularly centralising effect. Whereas a single agency maintaining their own dataset has the freedom to add in data fields or to restructure their working processes in order to meet a particular local need, when that data is managed as part of a centralised infrastructure, their ability to influence change in the way data is managed will be constrained both by the technical design and the institutional and funding arrangements of the data infrastructure (Miller & Roe, 2018). A more responsive government is not only about better intelligence at the centre, it is also about autonomy at the edges, and this is something that data infrastructures need to be explicitly designed to enable, and something that they are generally not oriented towards.

An analogy may be useful to illustrate the tensions present in how data infrastructure is designed. In the eighteenth century, Britain underwent an infrastructural transformation, with the development of a national network of roads and highways (Guldi, 2012). As metaled roads spread out across the country there were debates over whether to use local materials, which were easy to maintain with local knowledge, or to apply a centralised 'tarmacadam' standard to all roads. There were questions of how the network should balance the needs of the majority, with road access for those on the fringes of the country, and how the infrastructure should be funded. The shape, structure and design of this public infrastructure were highly contested, and the choices made over its design have had profound social consequences. Whilst Guldi uses the road network as an analogy for debates over modern internet infrastructures, it can also illuminate questions around equally intangible public data infrastructure.

If you build roads to connect the largest cities but leave out a smaller town, the relative access of people in that town to services, trade and wider society is diminished. In the same way, if your data infrastructure lacks the categories to describe the needs of a particular population, their needs are less likely to be met. Yet, a town might equally reject the idea of being connected directly to the road network, feeling that this might see its uniqueness and character eroded, much like some groups may also want to resist their categorisation and integration in the data infrastructure in ways that restrict their ability to self-define and develop autonomous solutions in the face of centralised data systems that are necessarily reductive (c.f. Rainie *et al.*, 2019).

Another debate that was central to the development of early transport infrastructure also resonates today for data infrastructure. That is the question of ownership, access and the public or private provision of infrastructure. Increasingly our nominally public data infrastructures may rely upon stocks and flows of data that are not publicly owned. In the United Kingdom, for example, the Postal Address File, which is the basis of any addressing service, was one of the assets transferred to the private sector when Royal Mail was sold off (Hope, 2013; Pollock & Lämmerhirt, 2019). The UK's national mapping agency, the Ordnance Survey, retains ownership and management of the Unique Property Reference Number (UPRN), a central part of the data infrastructure for local public service delivery, yet access to this is heavily restricted, and complex agreements govern the ability of even the public sector to use it (Ordnance Survey, 2016). Historically, authorities have faced major challenges in relation to 'derived data' from Ordnance Survey datasets, where the use of proprietary mapping products as a base layer when generating local records contaminates those local datasets with intellectual property

rights of the proprietary dataset, and restricts who they can be shared with (Yates *et al.*, 2018). Whilst open data advocacy has secured substantially increased access to many publicly owned datasets in recent years, when the datasets the state is using are privately owned in the first place, and only licensed to the state, the potential scope for public re-use and scrutiny of the data, and scrutiny of the policy made on the basis of it, is substantially limited.

In the case of smart cities this concern is likely to be particularly significant. Take transit data for example. In 2015 the City of Boston, Massachusetts, agreed a deal with Uber to allow access to data from the data-rich transportation firm to support urban planning and to identify approaches to regulation. Whilst the data shared revealed some information about travel times, the limited granularity rendered it practically useless for planning purposes. Boston turned to senate regulations to try to secure improved access to granular data (Vaccaro, 2016). Yet, even if the city does get improved access to data about movements via Uber and Lyft in the city, the ability of citizens to get involved in the conversations about policy from that data may be substantially limited by continued access restrictions on the data. With the introduction of privately owned sensors, networks and processes central to many smart city models, the extent to which the data infrastructure for public tasks cease to have the properties that we will shortly see are essential to a 'participatory' public data infrastructure is a question worth addressing.

Participatory public data infrastructure

Even if we assume that the growth of urban data infrastructures in the smart city is almost inevitable, the shape that data infrastructures will take is far from predetermined. For many decades, community activists, advocates and civic campaigners have sought to ensure that citizens have the right to shape their cities, creating a vibrant field of participatory practice. Participation practitioners have sought to empower residents in decisions over the physical architecture of the city, the design of public services and the distribution of infrastructure such as roads, bus services, schools and hospitals (see https://participedia.net/ for examples and methods). Participatory design of the data systems of the city has been less common. Yet, with the design of data infrastructure affecting the policy options open to the city, and affecting the distribution of power and autonomy within the city, it becomes vital to explore what it may take to create a participatory public data infrastructure.

For Gray (Gray, 2015; Gray & Davies, 2015; Gray & Venturini, 2015) who has done much to develop the concept of a participatory data infrastructure, it is crucial to secure wider public awareness and engagement with questions of how data are collected, structured and published. I suggest that in thinking about the participation of citizens in public data, there are three critical aspects:

- Participation in data use
- Participation in data production
- Participation in data design

These are different in kind, but any case of participation may also be different in degree. Arnstein's ladder of participation (1969) offers a useful analytical tool to understand that within any form of citizen engagement with data the extent of participation can range from tokenism through to full, shared decision making (Figure 7.1). And, as for all participation projects, any assessment of participation is not complete without an answer to the question 'who is participating?', and a consideration of how inclusive or exclusive the participation opportunities are.

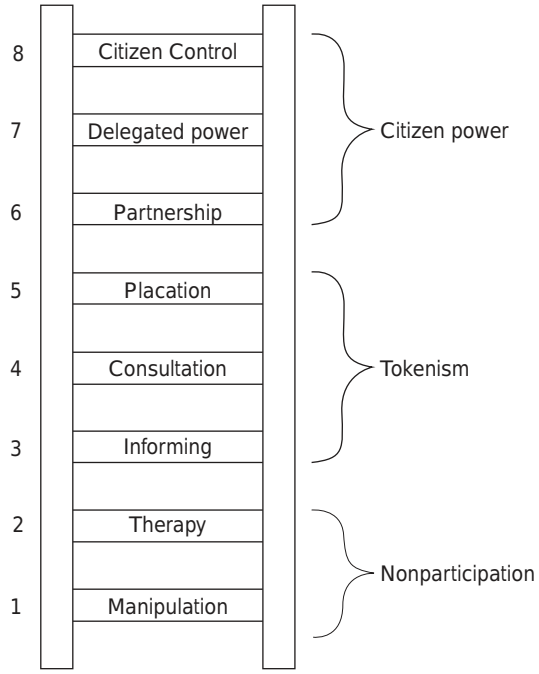

Figure 7.1 Arnstein's ladder of participation
Source: Arnstein (1969)

A data infrastructure assessed across all categories at the bottom-level 'non-participation' rung of Arnstein's ladder might record data 'about' citizens, without their active consent or involvement. It might exclude them from access to the data itself, and then uses the data to set rules, 'deliver' services and enact policies over which citizens have no influence in either their design or delivery. In such cases, the data infrastructure treats the citizen as an object, not an agent. Notably, this description might not be far off from the role that some fear is envisaged for citizens within a number of contemporary smart city visions (Krivý, 2016; Vanolo, 2016).

By contrast, when a data infrastructure is designed to enable citizens to participate in data use they are able to work with public data to engage in both service delivery and policy influence. Enabling this form of participation was a major focus for the early civic open data movement (Landry, 2019), drawing on ideas of co-production, and government-as-a-platform, to enable partnerships or citizen-controlled initiatives that made use of data to develop new solutions to local problems. In a more political sense, participation in data use can also address persistent patterns of information inequality between policy makers and citizens affected by policies, removing at least some of the power imbalances commonly at work when policies are discussed (Davies, 2010). In short, if the data relating to a city's population distribution and movement, electricity use, water connections, sanitation services and funding availability are shared, such that policy maker and citizen are working from the same data, then a data infrastructure can, in theory, act as an enabler of more meaningful participation, whether through new technology enabled channels or existing democratic fora.

Practical experience, however, often tells a different story. Instead of shared analysis of agreed data, it is not uncommon when engaging diverse citizen groups in seeking to use public data to find the process dominated by discussion of data gaps and ways in which the data itself needs to be improved. In some cases, the way data are being used might be uncontested, but the input might turn out to be misrepresenting the lived reality of citizens. This highlights the importance of citizen participation not only in data use, but also in dataset production. Simply having data collected from citizens does not make a data infrastructure participatory. The fact that sensors tracked my movement around an urban area does not make me an active participant in collecting data. By contrast, when citizens come together to collect new datasets, such as the water and air quality datasets generated by sensors from Public Lab (https://publiclab.org/wiki/nonprofit-initiatives), and when they are able to feed this into a shared corpus of data used by the state, there is much more genuine participation taking place. Similarly, the use of voluntary contributed data on Open Street Map, or submissions to issue-tracking platforms like FixMyStreet, can constitute a degree of participation in producing a public data infrastructure.

The opportunities for citizens are clearly positive when they use public data to shape policy and practice, and to contribute to the data on which policy and operational decisions are made. However, we should note that many participatory data projects, whether concerned with data use or with data production, are both patchy in their coverage and hard to sustain. They may offer an add-on to the public data infrastructure, but leave the core substantially untouched, not least because they generally rely on voluntary labour. This can also lead to substantial biases in terms of who gets to participate, what gets described and who benefits. Numerous studies have shown how contributions to crowd-sourced data platforms are often dominated by men, or by more affluent populations (Escher, 2011; Pak, 2017; Stephens, 2013), and thus tend to reflect their needs. There are also a limited number of topics that can attract a critical mass of user contributions adequate to address gaps in the state's own data infrastructure. For example, whilst crowdsourcing detailed urban cycle route data works in some cities, capturing detailed data on publicly accessible toilets with baby-changing facilities may be more difficult.

If then, participation is to have a sustainable impact on both policy and practice, it is important to consider how citizens can also be involved in shaping the core of public data infrastructure itself. This involves looking at the existing state-supported data collection activities that create public data infrastructure, and exploring whether or not choices over which data are collected, and how data are encoded, serve a broad public interest. If data collection and management practices are to allow the maximum range of choice and democratic freedom in policy making and implementation, then they cannot simply be taken as a given, but must themselves be subject to some degree of democratic oversight. This grounds a view of a participatory data infrastructure as one that enables citizens (and groups working on their behalf) not only to use and contribute data, but also to engage in discussions over data design.

The idea that communities and citizens should be involved in the design of data collection is not a new one. The history of public statistics owes a lot to the work of voluntary social reformers focused on health and social welfare in the eighteenth and nineteenth centuries, who initiated data collection to influence policy and then advocated for government to improve or take up on-going data collection (Dumpawar, 2015). In many countries, the design of the census, and of other government surveys, has long been a source of political contention (e.g. Thompson, 2010). Yet, with the vast expansion of connected data infrastructures, which rapidly become embedded, brittle and hard to change (Bowker & Star,

2000), we face a particular moment at which increased attention is needed to the place of citizens in shaping public data infrastructures. Seemingly technical choices, currently left to experts or to commercial providers, may set the long-term shape of public data infrastructures with substantial impact on the future of our urban environments.

It is the current expansion of data infrastructure that makes this a critical moment. Ribes and Baker (2007), in writing about the participation of social scientists in shaping research data infrastructures, underscore the importance of timing. They describe the limited window during which an infrastructure may be flexible enough to allow substantial insights from social science to be integrated into its development. As smart city agendas unfold, there may be a similarly limited window within which to establish data structures and systems that avoid foreclosing future political choices and that provide foundations for participatory public data infrastructure. Two critical tools to open and exploit this window come in the form of transparency policy and open data standard initiatives.

Part 2: Transparency and data standards

Although the movement for open data developed as a 'big tent', with different interests finding common ground around a limited technical definition of open data (Weinstein & Goldstein, 2012), arguments linking open data and transparency have played a particularly strong rhetorical role (Gray, 2014). Transparency was a central part of the framing of Obama's critical Open Government Directive in 2009, and transparency was in the foreground during the launch of data.gov.uk in the wake of a major political expenses scandal (Halonen, 2012). Transparency remains a core element of the narrative for many urban open data programmes.

It is also worth noting that transparency has also become an important resource in the regulatory toolbox of governments where governments mandate disclosure of information by private sector parties, for example, requiring publication of food safety inspection scores on restaurant windows and online (Fung *et al.*, 2007). As Fung et al. argue in *Full Disclosure*, governments have turned to these forms of targeted transparency as a way of requiring that certain information (including from the private sector) is placed in the public domain, with the goal of disciplining markets or influencing the operation of marketised public services by improving the availability of information upon which citizens will make choices (Fung *et al.*, 2007).

Thinking of open data as a transparency tool has two important consequences. Firstly, it draws a connection between open data as a form of proactive transparency, and pre-existing forms of reactive transparency, available through Right to Information laws. Sweden's Freedom of the Press law from 1766 was the first to establish a legal right to information (Michener, 2011), but it not until the middle of the last century that 'right to know' statutes started to appear elsewhere around the world. Today, over 100 countries have Right to Information laws in place, giving citizens a right to request documents held by the state. Increasingly, new or revised Right to Information laws recognise that transparency can be realised not only through access to documents, but also through providing access to datasets. Although there have been some tensions between document-based Right to Information communities, and more digitally oriented open data communities (Fumega, 2015; Janssen, 2012), there is a growing recognition of the complementary roles of reactive and proactive transparency through documents and data. When open data are not only provided by grant of government, but when citizens right to data is asserted, a key component of the right to the smart city is also enabled.

Secondly, linking open data and transparency provides an onwards link to accountability. Transparency policy is generally not only about access to information but also about asking

an actor to give account for their actions. Full transparency is only realised when information is received, processed and understood (Heald, 2006; Larsson, 1998). This creates space for citizens to not only access information, but also to demand that it is made comprehensible and usable as part of public scrutiny of decision making. This challenges 'data dumping' approaches to open data that rely on simply uploading some existing dataset to open data portals, and instead places the focus on the creation and publication of new more legible public datasets that respond to citizen demand.

It is unrealistic, however, to imagine that each city government will negotiate the shape, structure and contents of each open dataset to be disclosed with each different citizen who wants access. Rather, cities may be encouraged to adopt particular domain-specific transparency standards that can act to align their data publication with citizen needs. This makes standards themselves an important site for participation in the creation of urban data infrastructure.

Part 3: Standards

Data standards in the smart city come in many forms: from low-level specifications that describe the protocols through which sensor networks exchange packets of data, through to data schemas that set out the columns, fields, identifiers and business rules to use when structuring data for publication. In general, it is this latter form of standard that is brought into focus by transparency efforts. Transparency data standards function on the boundary between the internal data infrastructure of the state and the public realm.

The term standard itself has overlapping meaning within different communities of practice (Russell, 2014). For a technical community, a data standard specifies the 'how' of expressing certain data, describing the schema through which data may be represented. For a policy community, a data standard may be more about the 'what' of disclosure, setting out normative guidance on the particular information that should be provided. In practice, these levels of specification can interact, as the presence of a technical standard-as-schema offers the ability to automatically validate whether the particular information demanded by a standard-as-content-requirement has been provided. Standards may furthermore support the analysis of data in third-party tools that help render the data more directly legible as information for citizens and other stakeholders.

Although early open data advocacy focused on the idea of 'raw data now' (Davies & Frank, 2013), envisaging a one-way route from data inside government to re-use of that data by outside actors, as explored above the opening up of data can reveal the limitations of existing data infrastructure. Far from being a simple template into which data from internal systems is directly exported, the introduction of standards for publication of open data can lead to a reconfiguration of data practices within the state (Goeta & Davies, 2016), particularly when those standards are used to define what it means to be transparent in relation to a particular topic. This makes the design of standards, and the involvement of citizens in those design processes, an area of critical importance. Three brief examples may illustrate this point.

GTFS: defining transit transparency

GTFS stands for General Transit Feed Specification. It was originally developed in 2005 through collaboration between Google and staff at the City of Portland (McHugh, 2013). It has gone on to become the de-facto standard for disclosure of public transit schedules, providing the data to drive thousands of urban transit apps and underpinning smart city visions of more efficient urban mobility. Both private transit providers and government officials have had to

develop new working practices to turn schedules into GTFS format, and intermediary data infrastructures have been developed to bring together data from multiple existing datasets (Colpaert & Rojas Melendez, 2019; Goeta & Davies, 2016). However, although if you were to ask a city government if their public transport data were 'transparent' and they might point to their GTFS feeds, in practice these data only describe part of the transit story. For examples, fare information, transit performance or the accessibility of bus services is unlikely to be featured within the default GTFS datasets created. The data are tailored to real-time planning, but not to long-term transport planning. The 'official' version of the standard also lacks ways to describe informal transport models, particularly prevalent in the developing world, rendering the transport used by a large proportion of the population effectively invisible, unless governments adopt the independently maintained GTFS-flex extensions[1] (Colpaert & Rojas Melendez, 2019). The presence of a mechanism for independent extensions to GTFS does point, however, to a space for participatory engagement to at least partially reshape this component of smart city infrastructure. GTFS operates a loose governance model, in which a group of standard users is able to suggest and specify new elements that others are encouraged to use (https://developers.google.com/transit/gtfs/guides/changes-overview). Yet, unless these new elements feed into mainstream applications, or citizens campaign for their governments to publish data against them, they may remain theoretical rather than actual elements of government transparency.

OCDS: participation in design and adoption

When the Open Contracting Data Standard (OCDS) was developed in 2013, insights from GTFS adoption led the designers (the author of this article included) to include not only a technical specification, but also a set of data validation tools intended to support citizens to engage with their governments over the depth and breadth of information published about public procurement. OCDS emerged as the technical complement to a set of 'open contacting principles' (www.open-contracting.org/implement/global-principles/) that called on governments to publish detailed information about public tenders, contract awards and contracts signed with private parties (http://standard.open-contracting.org/). This information is particularly relevant in the context of the smart city, as smart city initiatives invoke Public Private Partnership (PPP) arrangements, license private vendors to collect data across the urban environment, or contract out delivery of public services. Fully implemented in a smart city, OCDS would allow citizens to understand and analyse this web of contractual relationships.

The development of OCDS followed an iterative process, starting from a review of supply-side data availability from government, and then mapping out a set of use-cases for transparent data on public procurement. By going back and forth between supply and demand, and consulting with potential data providers and users, the standard developed a set of data elements that struck a balance between the information citizens want and the data governments could theoretically produce. For example, a number of data use-cases called for geo-location of contracts, allowing public contracts and spending to be mapped by area. However, few, if any, e-procurement systems provide data fields to record location. As a result, the standard has specified location elements, but has not made these mandatory, providing space for citizens to engage with their own local governments to push for changes to upstream systems so that location information can be provided in future. This built on learning from Craviero et al.'s experience mapping public budget data in Sao Paulo, Brazil (Craveiro, 2013).

The governance and implementation model of OCDS provides two sites for participation public data infrastructure building. Firstly, the open governance process of the standard itself invites feedback from data users and publishers, and subjects proposed changes to a peer-review

process that invites representatives from government, civil society and private sector to approve or reject changes. Secondly, when any city sets out to adopt OCDS, implementation guidance encourages a dialogue with local civil society to identify priority use-cases for the data, and to use these to guide the prioritisation of changes to the internal data infrastructures that generate the data and the development of external data infrastructures for data analysis.

Air quality: in need of standardisation

Transparency data standards can facilitate not only citizen access to city data, but can also be the means by which citizens are engaged in the co-production of public data. Air pollution is an increasingly critical issue for cities to address, causing millions on premature deaths worldwide every year (Landrigan, 2017). Yet, as the Open Data Institute describe, 'we are still struggling to "see" air pollution in our everyday lives' (Fawcett, 2016). They report the case of decision-making around a new runway at Heathrow Airport, where policy makers were presented with data from just 14 'official' NO2 sensors. By contrast, a network of citizen sensors provided much more granular information, and information from citizen's gardens and households offered a contrasting account from to that presented with official sensor data.

Mapping data from official government air quality sensors in the UK reveals just how limited their coverage is and backs up the Open Data Institute's calls for a collaborative, or participatory, data infrastructure. Fawcett described how:

> Our current data infrastructure for air quality is fragmented. Projects each have their own goals and ambitions. Their sensor networks and data feeds often sit in silos, separated by technical choices, organizational ambition and disputes over data quality and sensor placement. The concerns might be valid, but they stand in the way of their common purpose, their common goals.
>
> *(2016)*

He concludes, 'We need to commit to providing real-time open data using open standards.'

This is a call for transparency targeted at both public and private actors: inviting those with access to sensors to allow re-use of their data, and to render it comparable with other data through use of common standards. The future design of such initiatives will need to carefully balance public and private interests, and to consider issues of privacy (Scassa, 2019), centralised or decentralised architectures of data exchange, and how to incentivise contributions from different stakeholders to a data commons, as well as the interaction between voluntary standardisation and data sharing imposed by regulation or via contracts. These are not only technical decisions, but decisions with substantial consequences for the kinds of public debates over air quality that can take place, and the way policy responses might be designed.

Part 4: Conclusion: transparency, standards and the shaping of infrastructure

In the preceding sections I have introduced the concept of open data and explained its broad based origins. I have suggested that the opening of city data supports a process of infrastructural inversion, in which the particular features of the underlying data infrastructure that produce it are brought into view and subject to debate. I have then argued that these data infrastructures are an important site for civic participation, not only through the use of data,

but also through citizen engagement in the (re-)design of the processes through which data are collected, structured and published. Without such participation, I argue, citizens become the object, rather than the subject, of the smart city. I have then outlined how the adoption of data standards to facilitate re-use of data creates a particular point for interventions that have the potential to shape the public data infrastructure that may result from transparency-driven open data initiatives.

Figure 7.2 offers a simplified schematic representation of the relationships at play, where processes of working with state data highlight limitations and surface new requirements, which can feed into standards and onwards into the internal data infrastructures of the state. The dotted line shows the boundary between data management 'inside' the state and the open engagement with data by citizens and other stakeholders in the lower part of the diagram.

It is important to note, however, that the 'virtuous cycle' described here, in which citizens are afforded space to participate in shaping public data infrastructures through data use and standardisation processes is by no means inevitable.

Firstly, a smart city initiative may lack a transparency-driven open data component. As Landry suggests (2019), although 'In some ways, open data is a golden thread running through modern urban development work' over recent years 'excitement for citizen–government collaboration based on a foundation of open data has often waned, and it is not clear how many cities have truly embedded a culture of openness through data into their organisational DNA'. Without both proactive and reactive open publication of datasets and an active citizenry requesting, exploring and engaging with data, opportunities to shape the data infrastructure of the smart city may be limited.

Secondly, there are competing pressures on a cities data infrastructure. Proprietary vendors may offer large 'integrated systems' in place of standards-based interoperable components. And even where standards are adopted, if these standards are proprietary or driven primarily by the

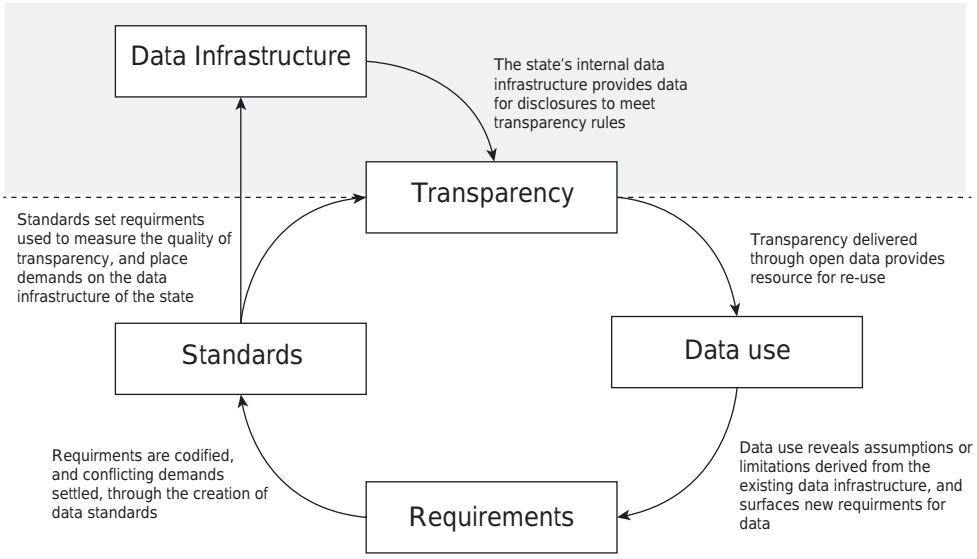

Figure 7.2 Mapping the relationship between transparency, data use, standards and data infrastructure

business requirements of commercial players in the smart city marketplace rather than by the concerns of citizens, infrastructures may be shaped to prioritise private, rather than public, interests. And even when citizens are engaged, if those citizens represent only a limited sub-set of the urban population, then the data infrastructures that emerge are likely to the tailored to their particular interests and needs of these niche groups. Just as good participation work involves proactive outreach to under-represented communities, the use of participatory spaces enabled by open data will generally require careful attention to be paid to who participates. Without this, gender and socio-economic biases often inherent within the technical communities are likely to come to the fore (Brandusescu & Nwakanma, 2019).

In short then, transparency, open data and standardisation are not automatic generators of participatory public data infrastructure. Instead, they offer a strategic space for pro-public interventions, and they call for new forms of practice from civic actors seeking to make sure the smart city is one which maximises, rather than restricts, the democratic freedoms of its residents. In closing then, let me offer six questions to support strategic engagement with the data infrastructure of your (smart) city:

(1) What information is pro-actively published, or can be demanded, as a result of transparency and right to information policies?
(2) What does the structure of the data reveal about the process/project it relates to?
(3) What standards might be used to publish such data?
(4) Do these standards provide the data that I, or other citizens, need to be empowered in relevant ways to this process/project?
(5) Are these open standards? Whose needs were they designed to serve?
(6) Can I influence these standards? Can I afford not to?

Note

1 https://github.com/MobilityData/gtfs-flex.

References

Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners*, 34(5), 216–224.

Bartenberger, M. & Grubmüller-régent, V. (2014). The enabling effects of open government data on collaborative governance in smart city contexts. *Journal of E-Decmoracy*, 6(1), 36–48.

Bartha, G. & Kocsis, S. (2011). Standardization of geographic data: The European inspire directive. *European Journal of Geography*, 2(2), 79–89.

Bowker, G. C. & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Brandusescu, A. & Nwakanma, N. (2019). Issues in open data: Gender equity. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (eds), *The state of open data: Histories and horizons* (pp. 287–299). Cape Town and Ottawa: African Minds and International Development Research Centre. www.stateofopendata.od4d.net/

Colpaert, P. & Rojas Melendez, J. A. (2019). Open data and transportation. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (eds), *The state of open data: Histories and horizons* (pp. 215–224). Cape Town and Ottawa: African Minds and International Development Research Centre. http://stateofopendata.od4d.net

Craveiro, G. da S., Santana, M. T. De, & Alburquerque, J. P. de. (2013). Assessing open government budgetary data in Brazil. In *ICDS 2013*. Retrieved from www.gpopai.usp.br/IMG/pdf/Craveiro-ICDS2013.pdf

Davies, T. (2010). Open data, democracy and public sector reform: A look at open government data use from data.gov.uk. Practical Participation. September 29. Retrieved from www.opendataimpacts.net/report/

Davies, T., Walker, S., Rubinstein, M., & Perini, F. (eds) (2019). *The state of open data: Histories and horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre. https://stateofopendata.od4d.net/

Davies, T. & Frank, M., 'There's no such thing as raw data': Exploring the socio-technical life of a government dataset, In *Proceedings of the 5th annual ACM web science conference*, pp. 75–78.

Denis, J. & Goëta, S. (2017). Rawification and the careful generation of open government data. *Social Studies of Science*, 47(5), 604–629.

Dodds, L. & Wells, P. (2019). Issues in open data: Data infrastructure. In T. Davies,S. Walker, M. Rubinstein, & F. Perini (eds), *The state of open data: Histories and horizons* (pp. 260–273). Cape Town and Ottawa: African Minds and International Development Research Centre. www.stateofopendata.od4d.net/

Dumpawar, S. (2015). *Open government data intermediaries: Mediating data to drive changes in the built environment*. Cambridge, MA: MIT Press.

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). *Digital era governance: IT corporations, the state, and E-government*. New York: Oxford University Press.

Escher, T. (2011). Writetothem. com: Analysis of users and usage for UK citizens online democracy. UK Citizens Online Democracy.

Fawcett, J. (2016) How to build the data infrastructure to tackle urban air pollution. https://oldsite.theodi.org/blog/how-to-build-the-data-infrastructure-to-tackle-urban-air-pollution

Floridi, L. (2004). Information. In L. Floridi (ed.), *The Blackwell guide to the philosophy of computing and information*. Wiley-Blackwell.

Fumega, S. (2015). Understanding two mechanisms for accessing government information and data around the world. World Wide Web Foundation, Washington, DC, July. https://idl-bnc-idrc.dspacedirect.org/handle/10625/56286

Fung, A., Graham, M., & Weil, D. (2007). *Full disclosure: The perils and promise of transparency* (1st ed.). Cambridge: Cambridge University Press.

Goeta, S. & Davies, T. (2016). The daily shaping of state transparency: Standards, machine-readability and the configuration of open government data policies. *Science & technology studies*, 29(4).

Goldsmith, S. & Crawford, S. (2014). *The responsive city*. John Wiley & Sons.

Gonzalez-Zapata, F. & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, 32(4), 441–452.

Government of Denmark. (2012). *Good basic data for everyone: A driver for growth and efficiency* (October). Copenhage: Government of Denmark.

Gray, J. (2014). *Towards a genealogy of open data*. Rochester, NY: Social Science Research Network 3 September. https://papers.ssrn.com/abstract=2605828

Gray, J. (2015). Democratising the data revolution: A discussion paper. Open Knowledge International.

Gray, J. & Davies, T. (2015). Fighting phantom firms in the UK: From opening up datasets to reshaping data infrastructures?. Working paper presented at the Open Data Research Symposium at the 3rd International Open Government Data Conference in Ottawa, on May 27th 2015.

Guldi, J. (2012). *Roads to power: Britain invents the infrastructure state*. Harvard University Press.

Gurin, J. (2014). Open Data Now. Open Data Now. McGraw-Hill Education.

Halonen, A. (2012). Being open about data analysis of the UK open data policies and applicability of open data. Report published by the Finish institute. www.fininst.uk/wp-content/uploads/2017/09/being-open-about-data.pdf

Heald, D. (2006). Varieties of transparency. *Proceedings of the British Academy*, 135, 25–43.

Hope, C. (2013). Everyone's postcodes to be privatised in Royal Mail flotation, despite objections from Sir Tim Berners-Lee. *The Telegraph*, 19 April. www.telegraph.co.uk/news/uknews/royal-mail/9994741/Everyones-postcodes-to-be-privatised-in-Royal-Mail-flotation-despite-objections-from-Sir-Tim-Berners-Lee.html

Huber, S. & Maier-rabler, U. (2012). The changing roles of citizens, civil society and public authority in Open Government. XXIInd World Congress of Political Science. Retrieved from http://paperroom.ipsa.org/papers/paper_11202.pdf

Janssen, K. (2012). Open government data and the right to Information: Opportunities and obstacles. *The Journal of Community Informatics*, 8(2).

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*. doi:10.1016/j.giq.2013.05.012

Kitchin, R. & Lauriault, T. P. (2014). Towards critical data studies : Charting and unpacking data assemblages and their work. In J. Eckert, A. Shears, & J. Thatcher (eds), *Geoweb and big data*. University of Nebraska Press. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112

Krivý, M. (2016). Towards a critique of cybernetic urbanism: The Smart City and the Society of Control. *Planning Theory*. doi:10.1177/1473095216645631

Landrigan, P. J. (2017). Air pollution and health, *The Lancet Public Health*, 2(1). e4–5. doi:10.1016/S2468-2667(16)30023-8

Landry, J. (2019). Open data and urban development. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (eds), *The state of open data: Histories and horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre. www.stateofopendata.od4d.net/

Larsson, T. (1998). How open can a government be? The Swedish experience. Openness and Transparency in the European Union. www.worldcat.org/title/openness-and-transparency-in-the-european-union/oclc/39554644

McHugh, B. (2013). Pioneering open data standards: The GTFS story. In B. Goldstein & L. Dyson (eds), *Beyond transparency: Open data and the future of civic innovation* (pp. 125–136). San Francisco: Code for America Press. http://beyondtransparency.org/pdf/BeyondTransparency.pdf#page=136

Michener, G. (2011). FOI laws around the world. *Journal of Democracy*, 22(2), 145–159.

Miller, D. & Roe, S. (2018). Registers and collaboration: Making lists we can trust. The Open Data Institute. https://theodi.org/article/registers-and-collaboration-making-lists-we-can-trust-report/

Ordnance Survey (2016). 'Policy statement – UPRN', Ordnance Survey. www.ordnancesurvey.co.uk/about/governance/policies/addressbase-uprn.html [accessed 28 June 2019].

Pak, B., Alvin, C., & Andrew Vande, M. (2017). FixMyStreet Brussels: Socio-demographic inequality in crowdsourced civic participation. *Journal of Urban Technology*, 24(2), 65–87. doi:10.1080/10630732.2016.1270047

Pollock, R. & Lämmerhirt, D. (2019). Open data around the world: European Union. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds), *The state of open data: Histories and horizons* (pp. 465–484). Cape Town and Ottawa: African Minds and International Development Research Centre. http://stateofopendata.od4d.net

Rainie, S.C., Kukutai, T., Walter, M., Figueroa-Rodríguez, O.L., Walker, J., & Axelsson, P. (2019). Issues in open data: Indigenous data sovereignty. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (eds), *The state of open data: Histories and horizons* (pp. 300–319). Cape Town and Ottawa: African Minds and International Development Research Centre. http://stateofopendata.od4d.net

Ribes, D. & Baker, K. (2007). Modes of social science engagement in community infrastructure design. *Proceedings of the 3rd Communities and Technologies Conference*, 107–130.

Russell, A. L. (2014). *Open standards and the digital age: History, ideology and networks*. New York: Cambridge University Press.

Sadoway, D. & Shekhar, S. (2014). (Re)Prioritizing citizens in smart cities governance: Examples of smart citizenship from urban india. *Journal of Community Informatics*, 10(3). http://ci-journal.org/index.php/ciej/article/view/1179

Scassa, T. (2019). Issues in open data: Privacy. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (eds), *The state of open data: Histories and horizons* (pp. 339–350). Cape Town and Ottawa: African Minds and International Development Research Centre. www.stateofopendata.od4d.net/

Sieber, R. E. & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, 32, 308–315. doi:10.1016/j.giq.2015.05.003

Star, S. L. & Ruhleder, K. (1996). Steps toward an Ecology of Infrastructure: Design and access for large information spaces. *Information Systems Research*, 17, 111–134.

Stephens, M. (2013). Gender and the GeoWeb: Divisions in the production of user-generated cartographic information. *GeoJournal*, 78(6), 981–996. doi:10.1007/s10708-013-9492-z

Tauberer, J. (2014). *Open government data: The book* (2nd ed.) https://opengovdata.io/

Thompson, D. (2010). 'The Politics of the census: Lessons from abroad', *Canadian Public Policy*. doi:10.3138/cp.36.3.377

US government (2009). Open Government Directive. Washington DC. Retrieved from www.whitehouse.gov/open/documents/open-government-directive

Vaccaro (2016) Boston wants better data from Uber, and is taking a roundabout route to try and get it. www.boston.com/news/business/2016/06/28/uber-data-boston-wants

Vanolo, A. (2016). Is there anybody out there? The place and role of citizens in tomorrow's smart cities. *Futures*, 82, 26–36. doi:10.1016/j.futures.2016.05.010

Weinstein, J. & Goldstein, J. (2012). The benefits of a big tent: Opening up government in developing countries. *UCLA Law Review Discourse*, 38, 38–48.

Yates, D., Keller, J., Wilson, R. & Dodds, L. (2018). The UK's geospatial data infrastructure: Challenges and Opportunities. The Open Data Institute. https://theodi.org/wp-content/uploads/2018/11/2018-11-ODI-Geospatial-data-infrastructure-paper.pdf/

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com