Government Information Quarterly 33 (2016) 472–480

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Privacy concerns in smart cities

Liesbet van Zoonen

Department of Sociology, Faculty of Social Sciences, Erasmus University Rotterdam, Netherlands

ARTICLE INFO

Article history:
Received 28 October 2015
Received in revised form 8 June 2016
Accepted 9 June 2016
Available online 1 July 2016

Keywords:
Privacy concerns
Smart city
City government
Big data
Open data

ABSTRACT

In this paper a framework is constructed to hypothesize if and how smart city technologies and urban big data produce privacy concerns among the people in these cities (as inhabitants, workers, visitors, and otherwise). The framework is built on the basis of two recurring dimensions in research about people's concerns about privacy: one dimensions represents that people perceive particular data as more personal and sensitive than others, the other dimension represents that people's privacy concerns differ according to the purpose for which data is collected, with the contrast between service and surveillance purposes most paramount. These two dimensions produce a 2×2 framework that hypothesizes which technologies and data-applications in smart cities are likely to raise people's privacy concerns, distinguishing between raising hardly any concern (impersonal data, service purpose), to raising controversy (personal data, surveillance purpose). Specific examples from the city of Rotterdam are used to further explore and illustrate the academic and practical usefulness of the framework. It is argued that the general hypothesis of the framework offers clear directions for further empirical research and theory building about privacy concerns in smart cities, and that it provides a sensitizing instrument for local governments to identify the absence, presence, or emergence of privacy concerns among their citizens.

1. Introduction and purpose

Local governments across the world are in the middle of technological and economic developments that come together in the catch-all label of 'smart cities'. In a smart city, ICT-infused infrastructures enable the extensive monitoring and steering of city maintenance, mobility, air and water quality, energy usage, visitor movements, neighborhood sentiment, and so on. Evidently, such processes use and produce massive amounts of data. In the Dutch city of Rotterdam, for instance, the traffic authority monitors about 22,000 vehicle movements every morning,[1] while the regional environment agency produces hourly data about air quality from sensors across greater Rotterdam resulting in over 175,000 observations per year. The promise of such large amounts of data for smarter management of cities extends to other sectors as well such as (predictive) policing, crowd control or public sentiment monitoring.

The notion of data, in this context, extends beyond the big numbers churned out by monitoring technologies, but also includes the data present in city registers, the data from government or corporate surveys and the data from social media updates. These data are ever more often combined and linked in order to produce joint indicators of city well-being, economic vitality or safety. Increasingly, local governments make these data also available to the wider public. All of this raises issues about who has legitimate access, which data can be opened up to public usage, what is the appropriate privacy framework for the linkage of different data? In these discussions, issues like political and public acceptance of smart cities are important as is the question of everyday experiences in such 'datafied' cities (Powell, 2014). While some claim that 'big data' will help cities become richer, cleaner and more efficient, others argue that cities will turn into data-driven robotic places where creativity and deviance have no place. Kitchin (2014a) argues that there is little attention for such 'politics of city data' nor for the question how particular practices of data collection and analysis may have problematic social effects. He adds that the ubiquitous collection of data about all city processes may produce 'panoptic' cities, in which "systems that seek to enable more effective modes of governance [may] also threaten to stifle rights to privacy, confidentiality, and freedom of expression" (p 12).

This paper forms part of these debates and starts from the assumption that it is necessary to acknowledge people's concerns about their privacy in the further development of smart cities in order to maintain their support and participation (e.g. Townsend, 2013), as will be explained in more detail in Section 3. Without such an acknowledgement smart city projects have been seen to become controversial and abolished.

E-mail address: Vanzoonen@fsw.eur.nl.

[1] http://www.ad.nl/ad/nl/1012/Nederland/article/detail/3668009/2014/06/06/Fileloze-vrijdag-groot-raadsel-net-zoveel-verkeer-als-anders.dhtml.

http://dx.doi.org/10.1016/j.giq.2016.06.004

The main purpose of the paper is to develop a framework for exploring people's specific privacy concerns in smart cities on the basis of existing research about people's privacy concerns in general. Two dimensions emerge from the literature: concerns differ with respect to the kind of data that are involved, which can range from personal to impersonal data and all degrees and combinations in-between; and concerns vary along the purpose for which purpose are data used, which can move from improving the livability and services in a city to advancing surveillance and keeping citizens in control. The data and purpose dimensions together form the basis for a 2 × 2 privacy framework in which smart technologies and applications as well as specific forms of data collection and usage can be plotted. The framework is then further explored and illustrated through a discussion of three concrete examples. It is argued that the framework offers an instrument for local governments to understand and incorporate privacy concerns in their policy and operational decisions, and that it offers a set of hypotheses to academic researchers to further conduct research about privacy concerns in smart cities.

2. Data landscapes in the smart city

Powell (2014) uses the term 'data cities' to indicate that 'smart' technologies like transport systems, air quality monitors or CCTV cameras simultaneously use and generate enormous amounts of data. Taylor and Richter (2015) similarly identify (big) data as key to the rise of smart cities. Bettencourt (2014) makes a more specific claim when she says that big data are particularly helpful for more successful urban policies and planning. Kontokoska (2015) speaks in this respect of computational urban planning. Both big data discourse and smart city discourse tend to obscure that data have always been crucial to city planning and city life. Cities started monitoring their populations mostly in the 19th century as part of the wider movement towards modern means of governing the nation state (e.g. Breckenridge & Szreter, 2012). Partly through civil registrations, partly through bureaucratic and commercial records, partly through surveys and mapping, the 19th century saw a similar data avalanche as we are witnessing today and data since has come to underpin city planning and decisions. Robertson and Travaglia (2015), therefore, claim that the current big data wave constitutes a difference of speed and size, but not one of analytic principle and relevance.

To date, such reflections on the historical and present-day importance of data for city management do not include a systematic inventory of the kinds of data involved. While it is not the purpose of this paper to provide such a catalogue, to understand the variety of privacy concerns at stake it is necessary to have at least a preliminary impression of the diversity of data that are used in and by cities. The table below presents such an impression for the city of Rotterdam in the Netherlands. It is based on the discussions, interactions and projects taking place in the Urban Big Data Lab, a collaboration of two Rotterdam universities and local government aimed at optimizing the understanding and usage of big, open and linked data for city policies and planning.[2]

While the above table is likely to be incomplete and imbalanced, it does convey the diversity of data in smart cities. Data differ in size, in regularity, in purpose, in complexity, in ownership, in visibility, and other matters. Moreover, within big cities oversight of these different data and streams tends to be lacking (cf. Meijer & Rodríguez Bolívar, 2015). City data emerges from a wide variety of governmental departments, from private and public stakeholders, from individual citizens and visitors, and are collected, analyzed and stored without any kind of central coordination or collaboration. Kaisler, Armour, Espinosa, and Money (2013) conclude that data diversifies and multiplies at unprecedented and unplanned speed, requiring ever bigger and multiple storage facilities and diverse and combined analytic techniques, while engaging different actors who tend to lack knowledge of each other let alone collaborate.

The complexity of the city data landscape has led many cities to appoint chief data officers who are responsible for the usage and management of data; in New York, in particular, a Mayor's Office of Data Analytics was established in 2013. Towns (2014, no page) provides the rationale for these decisions by saying that cities "have struggled to share and integrate data streams in ways that support comprehensive analysis. Issues around data ownership, as well as privacy laws and public perception, have been significant stumbling blocks."

3. Bringing citizens into the picture

The emerging city data landscape provides local governments with additional challenges as well. Al Nuaimi, Al Neyadi, Mohamed, and Al-Jaroodi (2015), for instance, identify five concrete, operational issues with respect to data, i.e. sources, sharing, quality, security, privacy and costs. (e.g. Kitchin, 2014a) takes a critical perspective and shows how the discourse around big data and smart cities produces a suggestion of data providing neutral information for rational governance, while hiding the political and corporate interests. In a similar vein, Söderström, Paasche, and Klauser (2014) analyze how the term 'smart city' has become a key theme in corporate storytelling, and argue for alternative understandings of smart cities that take public interests into account. Viitanen and Kingston (2014) provide a concrete analysis of problems that local governments face when confronted with a corporate push to adopt smart data technologies and big data applications, and show how there is a serious risk of following the imperatives of the market instead of the demands of public policy. According to Datta (2015), Kitchin (2014b) and Vanolo (2013), such public policy should, among other things, consider the uneven pace at which cities become smart. As is clear from the academic literature and even clearer from the explosion of conferences, seminars, networks, blog posts and social media updates, the development and the discourse around smart cities is carried by an urban 'tech-elite' of IT corporations, young, well-educated, mostly white and male professionals, and a-political aspiring city managers. In fact, anecdotal evidence suggests that the whole notion of 'smart city' or 'big data' and what it entails may be unknown to the majority of current city inhabitants and visitors (Thomas, Mullagh, Wang, & Dunn, 2015). Reflecting on Barcelona, March and Ribera-Fumaz (2014, p.1) argue therefore that it is imperative to "put citizens back at the center of urban debate".

Various suggestions have been made and explored to integrate a wider group of citizens into smart city design and policies, – for instance – through citizen participation (Berntzen & Johansson, 2016), crowd sourcing (Schuurman, Baccarne, De Marez, & Mechant, 2012), citizen-centered approaches (Gaved, Jones, Kukulska-Hulme, & Scanlon, 2012), or co-creation and living labs (Schaffers et al., 2011). Others have argued more generally for a stronger protection of the privacy of citizens living, working, shopping or travelling in a smart cit. Li, Dai, Ming, and Qiu (2015) identify the over-collection of data as a severe security risk, especially when it comes to the sensitive data that people hold on their smart phones. Martinez-Balleste, Perez-Martinez, and Solanas (2013) similarly fear for the privacy of citizens in smart cities, especially when it comes to protecting information about their identity, the kind of information they look for, their location, energy usage and possessions (see also Bartoli et al., 2011). Privacy scholars offering solutions to privacy risks in smart cities focus on particular technological solutions, such as cloud computing (Kahn, Pervez, & Ghafoor, 2014), privacy enhancing technologies (PETs, Rebollo-Monedero, Bartoli, Hernández-Serrano, Forné, & Soriano, 2014) or transparency enhancing technologies (TETs, Beran, Pignotti, & Edwards et al., no year). Policy makers have turned to privacy impact assessments as a tool to identify whether a specific technology or applications involves a privacy risk and how this can be mitigated (cf. Wright & De Hert, 2011).

[2] See http://www.kenniswerkplaats-urbanbigdata.nl/.

Table 1
City data landscape.

Sector	Domain	Kind of data	Example of application
Infrastructure	Transport and asset management, built environment	Monitoring data, registration data, geo data	Traffic and congestion patterns, real time dashboards
Sustainability	Energy usage, water, environment, weather	Sensor and monitoring data, civic measurement data	Air quality monitoring and pollution warnings
Health	Health, quality of life, well-being, life expectancy	Health data, survey data, lifelogging	Location specific noise levels and social or health problems in specific neighborhoods
Cohesion	Education, social capital, migration, neighborhoods, housing, crime	Survey data, civic and community web presence data	School quality in specific neighborhoods
Commerce	Business opportunities, marketing, location based services	Social media data, open government data	Investment maps for attracting new business
Experience	Events, leisure, nightlife, tourism, heritage	Social media data, archive data, sensor data	Real time social media analytics for crowd control

The problem with focusing on such technical or design solutions is that citizens themselves and their privacy concerns are not addressed. Yet, the scare research on actual citizen behavior in smart cities does suggest that the success of particular applications, such as smart cards, may depend more on citizens' perceptions of privacy and security risks than on the actual technological, design or policy guarantees of privacy (e.g. Belanche-Gracia, Casaló-Ariño, & Pérez-Rueda, 2015). Such a discrepancy between perceptions and official realities is somewhat similar to the situation with crime statistics. They are generally going down, while public perceptions of the risk of crime is going up, as – among others - British sociologist Furedi (2007) has discussed extensively. As a result, crime and safety policies are nowadays as much geared towards policing fear of crime, as it is to controlling crime itself (Scheider, Rowell, & Bezdikian, 2003). With privacy concerns in the smart city, the situation is likely to be the same: they need as much attention as the design of privacy itself. The following section, therefore, reviews people's privacy concerns.

4. Privacy concerns and privacy paradox

Privacy research has burgeoned in the past decades as a result of growing reliance of public and private institutions on digital interactions with citizens and consumers. Several national and international organisations have identified privacy as a key policy, regulatory and legislation challenge of the 21st century (e.g. Camenisch, Fischer-Hübner, & Rannenberg, 2011; Gartner, 2012; OECD, 2011). The research about people's privacy concerns is diverse and contradictory in terms of theory (Li, 2012, for instance, identifies 15 different theories of privacy in online contexts), methods (Van Zoonen, 2014 discuss the usage of experiments, surveys, qualitative interviews and document analysis) and outcomes (in particular with respect to the (lack of) influence of age, gender and other socio-demographic features on privacy concerns (ibid)). Moreover, the research has identified two important paradoxes. Despite people's clearly expressed concerns about their privacy, there is a simultaneous lack of appropriate secure behavior: the most popular pin code used is 1234[3] and many people use one password for multiple accounts (CIS, 2011; Van Zoonen, 2014). Moreover, they share their personal information on numerous social media sites, despite the fact that they do not feel very secure on, for instance, Facebook (Pew, 2015). In the relevant literature, this contrast between concerns and behavior is known as the 'privacy paradox' (e.g. Young & Quan-Haase, 2013). A further complication in privacy behavior comes from a 'control paradox' describing how the feeling of being in control over delivering or registering one's data leads to less concern about how one's data are later used by other parties (Brandimarte, Acquisti, & Loewenstein, 2013).

Notwithstanding the disorder in the field, three consistent dimensions come up over and again as factors influencing people's concerns about privacy: these relate to the type of data involved, the purpose of data collection and usage, and the organization or persons collecting and using the data.

4.1. Kinds of data

Although there are unequivocal legal definitions of what personal data or personally identifiable data are (e.g. in the General Data Protection Regulation of the EU and in US privacy law), people themselves have less consistent sensitivities when it comes to what they consider to be personal data. Various national and cross-national surveys have found that people consider medical, financial and civic data is considered highly sensitive, while one's nationality, gender or age are considered less problematic (a.o. BCG, 2013; Cranor, Reagle, & Ackerman, 2000; Eurobarometer, 2011; InfoSys, 2013). However, there is increasing anxiety about the possibility of combining seemingly impersonal data into highly personal citizen or consumer profiles (Harris, Sleight, & Webber, 2005; Tene & Polonetsky, 2012). When it comes to data emerging from biometric measurements there is variation in people's concerns: Prabhakar, Pankanti, and Jain (2003) claim that people find the usage of data from iris scans much less acceptable than usage of the data coming from systems of face recognition. People also differ in how sensitive they consider their social media updates or consumption patterns; for some people such data are highly private, for others they are trouble-free (Eurobarometer, 2011). While there is no research that has examined how people feel about the collection of impersonal data about, for instance, traffic flows or air quality, there is little reason to expect that people will be concerned about it. These data reveal nothing about individual people and hence will probably fall outside of the realm of privacy worries.

4.2. Purpose of data

Second, the research about privacy concerns suggests that people assess for which purpose data is used and weigh the benefits that providing their data may offer them. When these benefits are of immediate personal relevance (medical services, commercial gain), most people are willing to share their data with the organization asking for them (e.g. Acquisti, John, & Loewenstein, 2013). They do make, however, a tradeoff between the amount of data asked for and the benefits received in exchange and asking too much data quickly leads to a sense of being watched rather than being serviced (ibid). Benefits of data sharing that have a wider, social goal are less easily accepted as a worthwhile trade off. Sanquist, Mahy, and Morris (2008), for instance, found that the acceptance of the US government monitoring personal communications was high in the immediate aftermath of the 9/11 attacks but declined after about half a year.

A complicating factor for citizens and consumers to assess the purpose of data collection, comes from concern that their data are used

[3] http://www.telegraph.co.uk/technology/internet-security/10215714/Easy-as-1234-the-most-popular-PIN-codes-revealed.html.

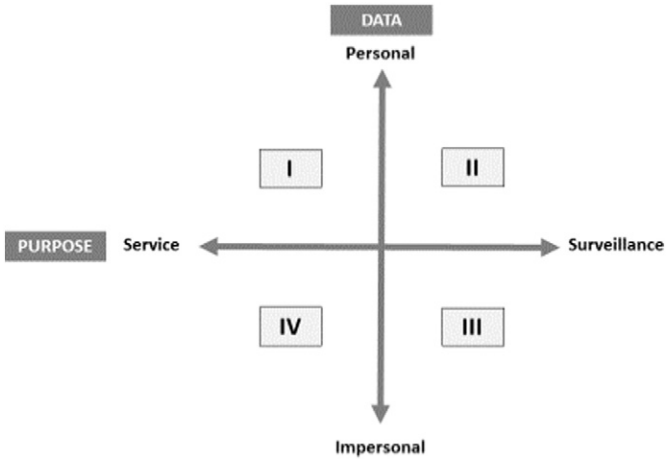

Fig. 1. Smart city privacy challenges.

for other purposes than they were originally collected for. While current EU regulation explicitly forbids this, it nevertheless keeps coming up as a major concern among citizens (Van Zoonen & Turner, 2013). Two examples further emphasize public suspicions about the secondary usage of their data: when, in The Netherlands in 2014, ING bank announced that it would share its client data with commercial parties, immediate public anger arose, people changed banks and in the end ING withdrew their plans and was forced to apologize (see Van Gaal, 2014). A data-sharing scheme of the UK National Health Service similarly came under attack when it appeared that the medical records kept by general practitioners would not only be shared with other health and care institutions, but also with commercial third parties, most notably health insurance companies (see Kirby, 2014). According to one UK newspaper, over 700.000 people chose to opt-out of the scheme as a result of the controversy (Dominiczak, 2015).

4.3. Who collects data?

People's concerns about their privacy also depend on who is dealing with their personal data. Eurobarometer data from 2011 indicate that across Europe people trust medical organisations and banks the most when it comes to their (most sensitive) data (Eurobarometer, 2011). On the lowest end of the spectrum, the Eurobarometer found telecom and internet companies, including social media and search engines; they are least trusted and sometimes even seen to threaten privacy (also BCG, 2013). In the context of this paper, local government is the key data-partner and hence the question is whether people trust local government with their data. In general, opinion polls consistently report that people tend to trust their local government much more than their national ones, for instance 72% trust in local government versus 24% in national government in the US (reported in Goldsmith, 2015), and 79% in local government versus 11% in national government in the UK (reported in Walker, 2013). Whether and to what extent citizens will trust their local governments to handle their personal data correctly and protect their privacy well, is as yet unclear. On the one hand, there is ample evidence of local government being an untrustworthy partner when it comes to data handling (Thompson, Ravindran, & Nicosia, 2015). On the other hand, there is an emerging debate whether opening up city data sets for usage by residents and other stakeholders will increase trust in local government; with one side of the argument being that "transparency and open data are important additions to the democratic toolbox" (O'Hara, 2012, no page), and the other warning against the myth that an open data policy is connected to transparency and more trust in government (Janssen, Charalabidis, & Zuiderwijk, 2012).

5. Privacy framework

The first two dimensions of people's privacy concerns, regarding the kinds of data and the purpose of data respectively, can be represented in a two-by-two scheme which identifies four types of possible sensitivities that people may have about smart city data. Given the absence of research about people's concerns about their local governments handling their data, we do not include this in the grid. The resulting figure is plotted in Fig. 1.

Because there is little concrete empirical research yet about how people experience their privacy in smart cities (but see Belanche-Gracia, Casaló-Ariño, & Pérez-Rueda (2015) for an exception), the actual placement of smart city technologies and data in one of the four quadrants in this framework is based on extrapolating and combining the research about people's privacy concerns discussed in the previous paragraph and the smart city data landscape suggested in Table 1.

5.1. Personal data used for service purposes (I)

All kinds of traditional data collected and registered by the city about its inhabitants are located in the first quadrant of the framework. It includes, for instance, city registrations like civil status (birth, death, and marriage), housing, elections or work. Data that people find more sensitive about their usage of social and economic care are also located in this quadrant. In the digital era, this quadrant has been expanded with data emerging from online transactions between city services and citizens, and from social media behavior of residents and visitors.

Local governments collect and use these data to monitor demographic patterns of its residents, assess the quality of its interactions with them and analyze civic moods. The purpose of these data is to underpin city management and planning, enhance city services and support local citizens. The privacy challenge in this quadrant is likely to be moderate, because, first, this kind of data has been part and parcel of city management all along and have rarely been subject to civic concerns. Second, because of the service purpose, citizens are likely to experience a positive tradeoff between handing over personal data and receiving, for instance, social benefits. However, especially in the latter case, there is a continuous risk of the perception of these data practices moving into the second quadrant, in which highly personal data are used for surveillance purposes. The Dutch System Risk Identification (SyRI), for instance, allows local governments to mine benefit registers in case of justifiable doubt about fraud. This is a controversial practice and particular groups of citizens may experience a thin line between service and surveillance here.

5.2. Personal data used for surveillance purposes (II)

This quadrant covers personal data, collected and monitored for surveillance purposes. It involves all police data, from minor violations to stop and search to criminal offence; it also involves data of local authorities such as public transport or port authorities. Digital and software innovations have added another layer to these data, for instance, the use of facial recognition software to analyze the images captured by CCTV cameras. Evidently, all such data are directly personal and citizens will often experience such data as highly sensitive (e.g. Samatas, 2008).

The combination of highly personal data collected and used for the purposes of surveillance and government control make this quadrant a highly contested one, with new data initiatives persistently under scrutiny of privacy advocates. The mayor of the French city Nice, for instance, won the ironic Big Brother Award in 2008 for his controversial decision to install the most pervasive and expensive video surveillance system in France. The city of Dresden received the award in 2012 for logging and tracing mobile phone traffic during a massive anti-Nazi demonstration.

Such infringes notwithstanding, the combination of personal data and surveillance purposes is subject to strong legal and regulatory

frameworks. The new EU General Data Protection Regulation, for instance, sets strict rules for the legitimate usage of personal data, offers a stronger position to citizens to control their data (including, among other things 'The right to be forgotten') and imposes high fines on data abuse, for which the data processor will be held responsible. For city governments, engaged in many new data initiatives, this means that a solid knowledge of the often complex privacy regulations becomes a new requirement as well (Cresswell & Pardo, 2002).

5.3. Impersonal data used for surveillance purposes (III)

The data in this quadrant concern all data that cannot be linked to an individual person and are used for surveillance and control purposes. Such data come from, for example, the monitoring of traffic flows, public transport, crowd, sports and event management through, among others, infrared video, CCTV or heat sensors. Such data may not automatically be perceived as sensitive, as they do not measure individuals but rather impersonal crowds or flows of vehicles. Similarly impersonal data for surveillance purposes comes from the aggregation and combination of survey and registration data in the city. The city of Rotterdam, for instance, has a Central Policy Information System in which data from various sources can be combined to answer pertinent or emerging policy questions. Such data are also often used in combination with geographical information systems. On the basis of postal codes and, for instance, police statistics, benefits and insolvency registers, housing and commerce information, particular city areas can be profiled as having high risks of economic or social unrest. Most developments around predictive policing and the identification of neighborhoods or streets with high crime risks, are based on the aggregation and articulation of data with postal codes (e.g. Perry, McInnis, Carter, Smith, & Hollywood, 2013).

All of such data can be analyzed and enhanced in ways that make it possible to identify individual citizens, thus making previous impersonal data suddenly highly personal. Through facial recognition software individuals in crowds can be recognized, and profiling on the basis of location data can be done so precisely that individual households can be identified. For such reasons, data practices like this have become subject to civil, political and individual suspicion, constituting a highly volatile policy arena. In the United States in particular, many civic organisations have protested against their local police acquiring predictive software, arguing that predictive algorithms inherit the biases present in standard policing, and perpetuate racist and prejudiced profiling (cf. Koss, 2015).

5.4. Impersonal data collected for service purposes (IV)

A big chunk of current smart city technologies and data usage concern impersonal data collected and used for the direct benefit of the city environment, the well-being of citizens and more efficient city operations. One can think of monitoring systems for air, noise and water quality; energy systems that are tailored to real-time usage; smart waste management; and so on. The data used for these applications are about 'things' rather than about people, and may therefore be less sensitive. All data that cities nowadays make available through their open data portals are also located in this quadrant, as privacy regulations prevent cities to publish any other kind of data (cf. Zuiderwijk, Janssen, Choenni, Meijer, & Alibaks, 2012). Note, however, that there are significant national differences with respect to open data, with Norway presenting an extreme case of making citizen tax returns available to all members of the public since 1800; currently the rank order of tax payers in Norway is online. Like in quadrant III, one finds here aggregated register and survey data in combination with location facts, producing information relevant to improve city services, rather than surveillance. City public health policies increasingly make use of such indicators, identifying, for instance, areas with high air or noise pollution and their correlation with particular disease patterns (e.g. Erdem, Prins, Voorham, Van Lenthe, and Burdorf, 2015).

The combination of impersonal data with service purposes seems to make this quadrant a rather 'innocent' one for policy and government, because security breaches and data-abuse are unlikely to have significant and direct effects on individual citizens. Nevertheless, here too, there are privacy concerns coming from increasingly detailed methods of profiling which may enable the re-identification of individuals from aggregate and anonymized data (e.g. Kitchin, 2014c).

6. Using the privacy framework

The privacy framework can be used in two combined ways: first, to develop a set of academic hypotheses that contribute to a more situated understanding of people's privacy concerns; and second, to understand the policy challenges that specific smart city technologies and data usage may throw up to local governments. Three examples will illustrate this double potential.

6.1. Example 1: smart waste technologies

Cities are currently all considering the acquisition of smart technologies for waste management to reduce cost and enhance efficiency. Several kinds of smart solutions have been introduced involving among other things, bins with sensors that measure how full the bin is. Only when a certain level of waste in the bin has been reached, a dustcart will come to empty it. According to some companies promoting their smart waste solutions, this can help save cities up to 50% of their costs in waste logistics.[4] Some types of smart bins only involve a sensor that measures the level of waste, whereas others enable the simultaneous authentication of the user through smart card access. When plotting these different solutions into the matrix, we see how the privacy concerns may change as well.

The bin in the below left quadrant of the matrix only measures if the bin is almost full and needs emptying. It is unclear who has thrown stuff in, but the system does make sure the bin will be emptied in time: impersonal data combined with a service purpose are unlikely to raise privacy concerns. The bin in the upper right quadrant requires card authentication by the person who wants to throw their garbage away. While authentication and the prevention of illegal dumping may be the main purpose of this system, it also enables the collection of highly personal data about who throws away how much in that particular bin. The additional data potential of the bin does comes with the cost of it moving from an innocent technology to a privacy sensitive one (Fig. 2).

The placement of the waste management technologies in the privacy framework is based on the extrapolation of the reviewed academic research about privacy concerns. When articulated more precisely in the context of city-waste management, the suggestion emerges that the choice and usage of a particular operational technology is a key variable for raising people's privacy concerns. This suggestion produces a useful hypothesis for research about people's privacy concerns regarding smart solutions for city problems, for instance those for smart parking (contrasting sensors indicating whether there is a car above the sensor with systems based on CCTV registrations); or for smart lighting (contrasting measuring people movement by heat sensors with systems using cameras).

6.2. Example 2: predictive policing

Predictive policing involves using individual and aggregate data to analyses crime patterns and enhance police performance. Initially, predictive policing took place on the basis of statistics of the places and types where particular crimes happened in the past: thus, areas with high reports of burglary and vehicle theft were identified and expected

[4] E.g. http://www.enevo.com/wp/wp-content/uploads/2013/05/a4_enevo_general_en_v4.pdf; or http://www.urbiotica.com/en/smart-solutions/intelligent-waste-management/.

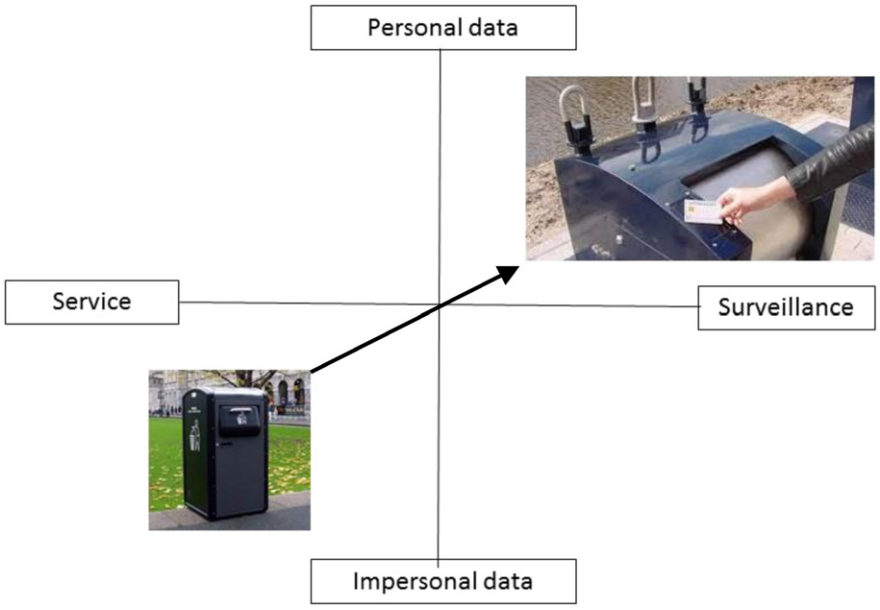

Fig. 2. Contrasting privacy concerns for smart waste management.

to be subject to more crime in the future. Additional police patrol in these areas during the predicted time slots, is said to have significantly reduced those crimes in a number of US cities (Braga, Papachristos, & Hureau, 2014). The data used for such forms of predictive policing are impersonal and aggregated; hence such form of predictive policing can be located in the lower right quadrant of the smart city privacy challenges. However, predictive policing may move quickly to the upper right quadrant of the model if the data mining techniques are used beyond the identification of times and places, and try to predict not only types of crime but also types of offenders. Based on data mining of the features of convicted criminals, risk profiles are made which can identify who is likely to commit a crime in the near future. Increasingly, social media data come into the equation and it is the particular linking of diverse data sets which changes the practice of preventive policing into a highly problematic one that touches immediately on the privacy of individuals. Privacy advocates argue that the predicted combination

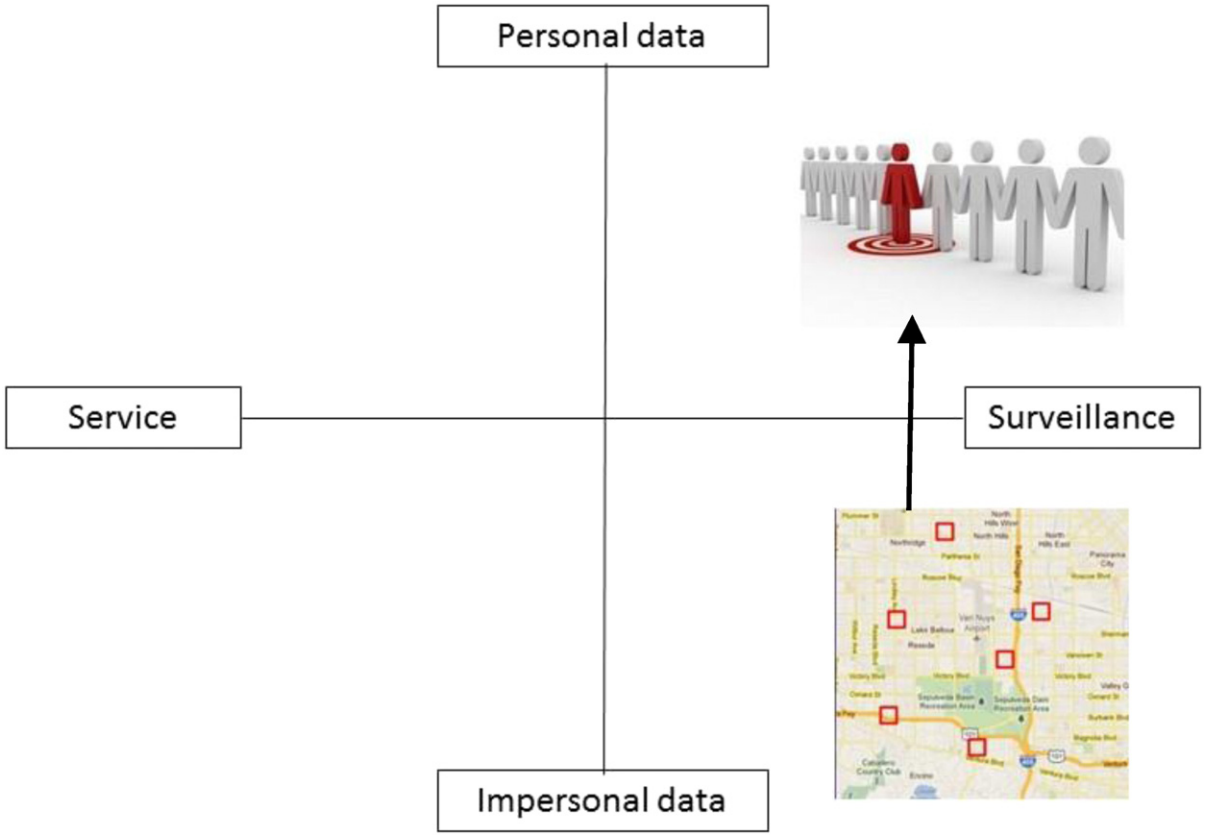

Fig. 3. Contrasting governance challenges for predictive policing.

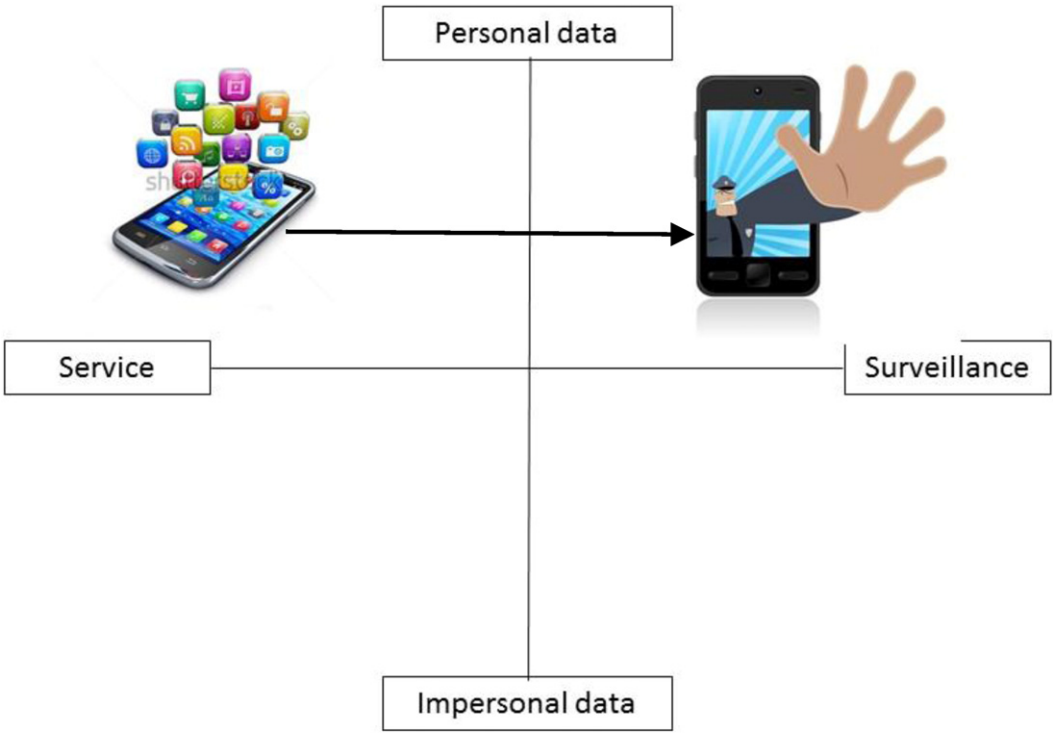

Fig. 4. Contrasting governance challenges for social city media.

of place and types of people turns all inhabitants of a particular city into potential criminals.

Here too, the placement of predictive policing in the privacy framework is based on the extrapolation of the reviewed academic research about privacy concerns. The hypothesis that can be developed from this example is that privacy concerns will vary with the particular (combinations of) data and analytic tools that are used, contrasting, not only, aggregated procedures to the ones that enable individual profiling, but also contrasting service to surveillance purposes. Does individual profiling raise less concern if people perceive they profit immediately and tangibly from enhanced city services, a question that is elaborated in the next example (Fig. 3).

6.3. Example 3: social media monitoring

Most cities nowadays use an array of social media to promote themselves and communicate with their citizens. The city of Rotterdam, for instance, has more than 30 thematic and neighborhood Twitter accounts, several Facebook and LinkedIn pages, and furthermore an extensive presence on YouTube, Flickr, Instagram and Pinterest. The city has been praised for its activities and for its quick response to citizen requests living up to its rule that every request should have a response within two hours.[5] The data generated on these social media accounts are systematically monitored through dedicated social media analytics and sentiment analysis, and provide information about the particular networks that the Rotterdam civil servants are engaged in, but also in the networks of Rotterdam inhabitants, public opinion, tensions in particular areas, and so on. The data produced and collected through social media are by definition tied to individuals and hence personal, although some people do not consider their social media behavior as personal (as discussed in Section 4.1). The purpose of social media usage by local governments is mostly to provide services and enhance the city's responsiveness to its citizens. In addition, there is an increasing interest to explore whether social media analytics can replace more extensive data sources such as a yearly or bi-annual city survey or census (Data for Policy, 2015).

It is thus firmly located in the upper left quadrant of the model. However, as with other forms of social media analytics, the potential to mine the data for more personalized information and targeting of city services are endless and hence, there is a continuous risk of these services being pulled towards the more problematic quadrant where privacy is at stake, and purpose may shift away from service to surveillance.

Like in the example about predictive policing, the main hypothesis to be examined in more detail here, is that privacy concerns will depend on the (combinations of) data, and analytic tools and procedures that are used to enhance city services. Moreover, given the research about costs and benefits of revealing personal data (discussed in Section 4), an additional question to explore is this depends on the kind of services the city offers (Fig. 4).

7. Conclusion and implications

In this paper a framework was presented to identify what kind of privacy concerns the use of smart technologies and of (big, open and linked) data produce may raise among people (as citizens, workers, consumers or travelers) in smart cities Asserting, on the basis of existing research about privacy perceptions, that these concerns are underpinned by the way people perceive particular data as personal or impersonal, and that their concerns differ according to the purpose for which data is collected (service or surveillance), four areas of concerns emerge that range from hardly any (impersonal data, service purpose), to extremely high (personal data, surveillance purpose). The framework was then used to explore how specific technologies (smart bin, smart parking), and data usage (predictive policing, social media monitoring) may produce variable privacy concerns. This resulted in a general hypothesis that the choice of smart technologies and the usage of particular (combinations of) data and analytic tools are crucial factors to understand people's privacy concerns in smart cities (in addition to their perception of what kind of data for which purpose are being

[5] http://www.socialmediameetlat.nl/pdf/digiloog/36%20berrevoets.pdf.

used). There is both an applied need to further substantiate the empirical relation between data, purpose and technology/tools, and a fundamental one to produce a more situated, theoretical understanding of people's privacy concerns in smart cities. Using the framework will make the systematic accumulation of such concrete investigations easier, but also enables consistent comparisons between cities, in national as well as international contexts.

The hypothetical status of the framework does not prevent its usability; it can helpfully operate as a sensitizing instrument for policymakers and operational managers in the smart city flagging up in which contexts privacy concerns among their citizens may occur. Some of these concerns are covered by the new EU regulation regarding the processing of personal data. However, as it becomes ever easier to construct individual profiles from impersonal data, additional strategies and regulation may be necessary (cf. Data for Policy, 2015). Moreover, smart city technologies and data developments are so quick and ubiquitous that official legislation may fall short for the decades to come. Finally, as the earlier sections about the privacy paradox showed (see Section 4), people's concerns and perceptions are not always very consistent or predictable. For smart city governments, the challenge is thus threefold:

- identify which privacy concerns for their citizens may be at stake with specific technologies and data practices;
- identify if and how these are subject to the EU data protection regulation;
- develop a specific city policy on new developments that accommodates the concerns of citizens, beyond the bare legal necessities.

While these recommendations may seem too obvious, individual citizens or collective citizen groups are often ignored as partners in the development of smart city technologies or innovations (cf. Leydesdorff, 2012). However, the input and support of individual and collective civil actors is of crucial importance, as they will have to live with and within the smart datafied cities on an everyday basis.

References

Acquisti, A., John, & Loewenstein, G. (2013). What is privacy worth? *The Journal of Legal Studies*, *42*(2), 249–274.

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, *6*(1), 1–15.

Bartoli, A., Hernández-Serrano, J., Soriano, M., Dohler, M., Kountouris, A., & Barthel, D. (2011, December). Security and privacy in your smart city. *Proceedings of the Barcelona smart cities congress*.

BCG (2013). The value of our digital identity. *Liberty global policy series*Boston Consultancy Group.

Belanche-Gracia, D., Casaló-Ariño, L. V., & Pérez-Rueda, A. (2015). Determinants of multi-service smartcard success for smart cities development: A study based on citizens' privacy and security perceptions. *Government Information Quarterly*, *32*(2), 154–163.

Beran, S., Pignotti, E., & Edwards, P. (d). Trusted tiny things: Making devices in smart cities more transparent. http://ceur-ws.org/Vol-1280/paper9.pdf (no year, last accessed January 27, 2016)

Berntzen, L., & Johansson, M. R. (2016). The role of citizen participation in municipal Smart City projects: Lessons learned from Norway. *Smarter as the new urban agenda* (pp. 299–314). Springer International Publishing.

Bettencourt, L. M. (2014). The uses of big data in cities. *Big Data*, *2*(1), 12–22.

Braga, A. A., Papachristos, A. V., & Hureau, D. M. (2014). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Quarterly*, *31*(4), 633–663.

Brandimarte, L., Acquisti, A., & Loewenstein, G. (2013). Misplaced confidences privacy and the control paradox. *Social Psychological and Personality Science*, *4*(3), 340–347.

Breckenridge, K., & Szreter, S. (2012). *Registration and recognition: Documenting the person in world history*. Oxford University Press.

Camenisch, J., Fischer-Hübner, S., & Rannenberg, K. (Eds.). (2011). *Privacy and identity management for life*. New York/Heidelberg: Springer Press.

CIS (2011). Password security; a survey of Australian attitudes towards password user management. Centre for Internet Safety with PayPal. https://www.paypal-media.com/assets/pdf/fact_sheet/cis_paypal_whitepaper_final.pdf

Cranor, L. F., Reagle, J., & Ackerman, M. S. (2000). *Beyond concern: Understanding net users' attitudes about online privacy*. Cambridge, MA: MIT Press, 47–70.

Cresswell, A. M., & Pardo, T. A. (2002). Implications of legal and organizational issues for urban digital government development. *Government Information Quarterly*, *18*(4), 269–278.

Data for Policy (2015). A study of big data and other innovative data-driven approaches for evidence-informed policymaking. *Report of the international workshop on data driven innovations for better policies, 22 September 2015 in Brussels* (http://media.wix.com/ugd/c04ef4_29913c6b4f1341b8b6faffb520eefadd.pdf, last accessed October 28, 2015).

Datta, A. (2015). A 100 smart cities, a 100 utopias. *Dialogues in Human Geography*, *5*(1), 49–53.

Dominiczak, P. (2015). Nearly one million patients could be having confidential data shared against their wishes. The telegraph, June 5. http://www.telegraph.co.uk/news/health/news/11655777/Nearly-1million-patients-could-be-having-confidential-data-shared-against-their-wishes.html (Last accessed, October 24, 2015)

Erdem, Ö., Prins, R., Voorham, T., Van Lenthe, F., & Burdorf, A. (2015). Structural neighbourhood conditions, social cohesion and psychological distress in the Netherlands. *The European Journal of Public Health*. http://dx.doi.org/10.1093/eurpub (2015).

Eurobarometer (2011). Attitudes on data protection and electronic identity in the European Union. *Special barometer 359*Brussels: European Commission, Directorate-General for Communication.

Furedi, F. (2007). The only thing we have to fear is the 'culture of fear' itself. *American Journal of Sociology*, *32*, 231–234.

Gartner (2012). The future of the internet. Fundamental trends, scenarios and implications to heed. https://www.gartner.com/doc/2119220?ref=SiteSearch&sthkw=identity%20management&fnl=search (last accessed July 14, 2014)

Gaved, M., Jones, A., Kukulska-Hulme, A., & Scanlon, E. (2012). A citizen-centred approach to education in the smart city: Incidental language learning for supporting the inclusion of recent migrants. *International Journal of Digital Literacy and Digital Competence*, *3*(4), 50–64.

Goldsmith, R. (2015). Why trusts in local governments should be even higher than it is. Governing, the states and localities, August 18. http://www.telegraph.co.uk/news/health/news/11655777/Nearly-1million-patients-could-be-having-confidential-data-shared-against-their-wishes.html (last accessed October 24, 2015)

Harris, R., Sleight, P., & Webber, R. (2005). *Geodemographics, GIS and neighbourhood targeting, Vol. 7*, John Wiley and Sons.

Infosys (2013). Engaging with digital consumers. They are ready, are you? http://www.infosys.com/marcom/digital-consumer-study/default.asp (last accessed July 14, 2014)

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, *29*(4), 258–268.

Kahn, Z., Pervez, Z., & Ghafoor, A. (2014, December). Towards cloud based smart cities data security and privacy management. *Utility and cloud computing (UCC), 2014 IEEE/ACM 7th international conference on* (pp. 806–811). IEEE.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. *System sciences (HICSS), 2013 46th Hawaii international conference on* (pp. 995–1004). IEEE.

Kirby, T. (2014). Controversy surrounds England's new NHS database. *The Lancet*, *383*(9918), 681.

Kitchin, R. (2014a). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14.

Kitchin, R. (2014b). Making sense of smart cities: Addressing present shortcomings. *Cambridge Journal of Regions, Economy and Society*, *8*, 131–136.

Kitchin, R. (2014c). *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage.

Kontokoska, C. E. (2015). *The quantified community and neighborhood labs: A framework for computational urban planning and civic technology innovation.* (Available at SSRN 2659896).

Koss, K. K. (2015). Leveraging predictive policing algorithms to restore fourth amendment protections in high-crime areas in a post-Wardlow world. *Chicago-Kent Law Review*, *90*(1), 301–334.

Leydesdorff, L. (2012). The triple helix, quadruple helix, …, and an N-tuple of helices: Explanatory models for analyzing the knowledge-based economy? *Journal of the Knowledge Economy*, *3*(1), 25–35.

Li, Y. (2012). Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems*, *54*(1), 471–481.

Li, Y., Dai, W., Ming, Z., & Qiu, M. (2015). *Privacy protection for preventing data over-collection in smart city*.

March, H., & Ribera-Fumaz, R. (2014). Smart contradictions: The politics of making Barcelona a self-sufficient city. *European Urban and Regional Studies* (0969776414554488).

Martinez-Balleste, A., Perez-Martinez, P., & Solanas, A. (2013). The pursuit of citizens' privacy: A privacy-aware smart city is possible. *IEEE Communications Magazine*, *51*(6), 136–141.

Meijer, A., & Rodríguez Bolívar, M. P. (2015). Governing the smart city: A review of the literature on smart urban governance. *International Review of Administrative Sciences*. http://dx.doi.org/10.1177/0020852314564308 (first published online on April 29, 2015).

OECD (2011). Digital identity management. Enabling innovation and trust in the internet economy. (Available at) http://www.oecd.org/internet/interneteconomy/49338380.pdf (last accessed December 12, 2012)

O'Hara, K. (2012). Transparency, open data and trust in government: Shaping the infosphere. *Proceedings of the 4th annual ACM web science conference* (pp. 223–232). ACM.

Perry, W. L., McInnis, B., Carter, C., Smith, S., & Hollywood, J. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. Washington: Rand Corporation.

PEW (2015). Americans' privacy strategies post-Snowden. http://www.pewinternet.org/files/2015/03/PI_AmericansPrivacyStrategies_0316151.pdf

Powell, A. (2014). 'Datafication', transparency, and good governance of the Data City. *Digital enlightenment yearbook 2014: Social networks and social machines, surveillance and empowerment* (pp. 215).

Prabhakar, S., Pankanti, S., & Jain, A. K. (2003). Biometric recognition: Security and privacy concerns. *IEEE Security and Privacy*, *2*, 33–42.
Rebollo-Monedero, D., Bartoli, A., Hernández-Serrano, J., Forné, J., & Soriano, M. (2014). Reconciling privacy and efficient utility management in smart cities. *Transactions on Emerging Telecommunications Technologies*, *25*(1), 94–108.
Robertson, H., & Travaglia, J. (2015). Big data problems we face today can be traced to the social ordering practices of the 19th century. *The impact blog*. London School of Economics (http://blogs.lse.ac.uk/impactofsocialsciences/2015/10/13/ideological-inheritances-in-the-data-revolution/, last accessed January 27, 2016).
Samatas, M. (2008). From thought control to traffic control: CCTV politics of expansion and resistance in post-Olympics Greece. *Sociology of Crime Law and Deviance*, *10*, 345–369.
Sanquist, T. F., Mahy, H., & Morris, F. (2008). An exploratory risk perception study of attitudes toward homeland security systems. *Risk Analysis*, *28*(4), 1125–1133.
Schaffers, H., Sällström, A., Pallot, M., Hernández-Muñoz, J. M., Santoro, R., & Trousse, B. (2011, June). Integrating living labs with future internet experimental platforms for co-creating services within smart cities. *Concurrent enterprising (ICE), 2011 17th international conference on* (pp. 1–11). IEEE.
Scheider, M. C., Rowell, T., & Bezdikian, V. (2003). The impact of citizen perceptions of community policing on fear of crime: Findings from twelve cities. *Police Quarterly*, *6*(4), 363–386.
Schuurman, D., Baccarne, B., De Marez, L., & Mechant, P. (2012). Smart ideas for smart cities: Investigating crowdsourcing for generating and selecting ideas for ICT innovation in a city context. *Journal of theoretical and applied electronic commerce research*, *7*(3), 49–62.
Söderström, O., Paasche, T., & Klauser, F. (2014). Smart cities as corporate storytelling. *City*, *18*(3), 307–320.
Taylor, L., & Richter, C. (2015). Big data and urban governance. *Geographies of urban governance* (pp. 175–191). Springer International Publishing.
Tene, O., & Polonetsky, J. (2012). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property*, *11*(5) (online).
Thomas, V., Mullagh, L., Wang, D., & Dunn, N. (2015). Where's Wally? In search of citizen perspectives on the smart city. *8th conference of the international forum on urbanism (IFoU)* (pp. 1–8). Multidisciplinary Digital Publishing Institute.
Thompson, N., Ravindran, R., & Nicosia, S. (2015). Government data does not mean data governance: Lessons learned from a public sector application audit. *Government Information Quarterly*, *32*(3), 316–322.
Towns, S. (2014). Which states and cities have chief data officers? Government Technology, June 23. http://www.govtech.com/state/Which-States-and-Cities-Have-Chief-Data-Officers.html (last accessed October 23, 2015)
Townsend, A. (2013). *Smart cities. Big data, civic hackers and a quest for a new utopia*. W.W. Norton Company.
Van Gaal, M. (2014). ING plan to share customer payment data spurs privacy concerns. http://www.bloomberg.com/news/articles/2014-03-10/ing-plan-to-share-customer-payment-data-spurs-privacy-concerns (Last accessed, October 24, 2015)
Van Zoonen, L. (2014). What do users want from their future means of identity management? Final report. http://imprintsfutures.org/assets/images/pdfs/End%20report%20IMPRINTS.pdf
Van Zoonen, L., & Turner, G. (2013). Taboos and desires of the UK public for identity management in the future: Findings from two survey games. *ACM digital identity management workshop, Germany, 8th November 2013*.
Vanolo, A. (2013). Smart mentality: The smart city as disciplinary strategy. *Urban Studies*, *51*(5), 883–898.
Viitanen, J., & Kingston, R. (2014). Smart cities and green growth: Outsourcing democratic and environmental resilience to the global technology sector. *Environment & Planning A*, *46*, 803–819.
Walker, D. (2013). Good people trust local government, but councils need to trust people. The Guardian, March 1. http://www.theguardian.com/local-government-network/2013/mar/01/people-trust-local-government (last accessed October 24, 2015)
Wright, D., & De Hert, P. (Eds.). (2011). *Privacy impact assessment*. Springer Science & Business Media.
Young, A. L., & Quan-Haase, A. (2013). Privacy protection strategies on Facebook: The internet privacy paradox revisited. *Information, Communication & Society*, *16*(4), 479–500.
Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-technical impediments of open data. *Electronic Journal of e-Government*, *10*(2), 156–172.

Liesbet van Zoonen is professor of Sociology at Erasmus University Rotterdam, and Academic Director of the Leiden-Delft-Erasmus Centre for Big Open and Linked Data Cities (www.boldcities.nl).
She is also the co-founder of the Knowledge Lab Urban Big Data, a collaboration between the city of Rotterdam, Erasmus University Rotterdam, and the Rotterdam University of Applied Sciences.
Her background is in political science, communication studies and sociology.