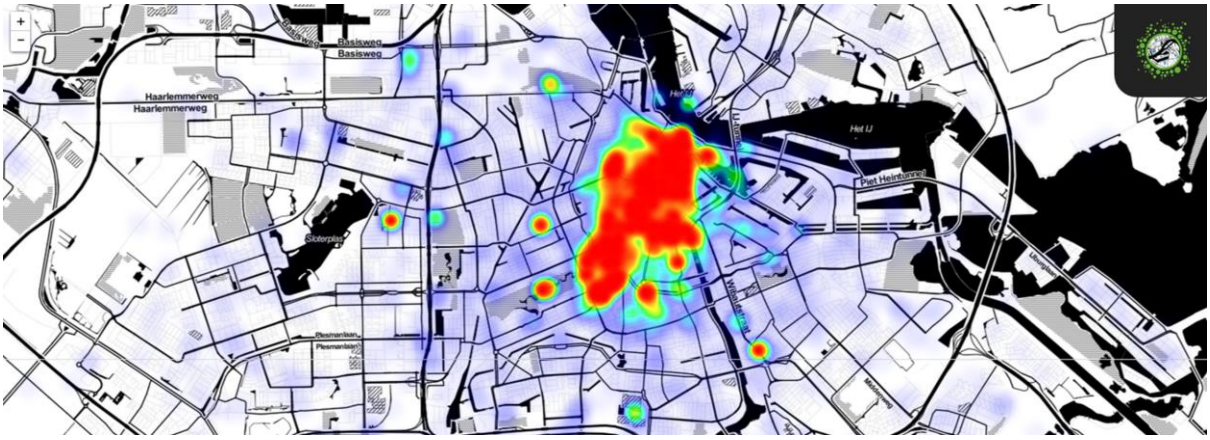
Customers, users or citizens?
Inclusion, spatial data and governance in the smart city

A report by Linnet Taylor, Christine Richter, Shazade Jameson and Carmen Perez del Pulgar

University of Amsterdam

Governance and Inclusive Development research group

With input from Professor Isa Baud, Dr. Karin Pfeffer, Alexandra Ruina, Christopher Livett, Nathalia Vredeveld and Dionne Poulussen



SocialGlass

maps4society

University of Amsterdam

This report is based on a grant made by Maps4Society (KIP 13759). The research was conducted at the University of Amsterdam in the department of International Development Studies, within the Governance and Inclusive Development research group. Co-Principal Investigators were Prof. Isa Baud and Dr. Linnet Taylor.

Executive Summary

Introduction

This report is based on a year of research into **how citizens in Amsterdam are becoming producers of digital data through their use of technology**, and the ways in which that data is becoming – or will likely become in the future – part of the way the city is governed. We focused primarily on spatial data (geo-information), defined as any digital data that indicates a person's location or movements. Today, we produce spatial data with everything we do, and in the future, it is likely that these data generated by city infrastructure and registration systems will become merged and linked with data generated directly by city residents such as social media postings, data from self-tracking devices and smart homes, maps generated by crowdsourcing, and feedback of all kinds.

We have a specific focus on the citizen's perspective. What kind of governance of digital data creates an equal playing field for the elderly, the young, the vulnerable or marginalised? For non-users of smart technologies, non-citizens, speakers of other languages? With this in mind, we conducted 20 expert interviews and 8 focus groups in Amsterdam over the course of 2015, aiming to include participants who were currently missing from, or marginalised by, current discussions and practices of smart city development, and also those whose lives might be changed most by an increase in urban datafication. Our discussions highlighted several groups: non-natives; ethnic or religious minorities; children and the elderly; those who opted out of using the technologies currently seen as necessary for citizen involvement in the smart city (i.e. smartphones); those who operate in highly regulated professions, and freelancers who are responsible for their own working environment.

Future scenarios

Based on our expert interviews we constructed four possible future scenarios which then informed the focus group discussions. The first, **data utopia/dystopia**, combines a situation where individuals are highly traceable with strong city control over data, monitoring both public and private spaces in real time. This means the city can profile people in great detail and to target policies and services to a neighbourhood or even household level, leading to efficient service provision and control over public safety – but it also leads to social engineering by policymakers and researchers, and tension that ultimately decreases social cohesion. The second scenario, **'Anonydam'**, involves greater individual anonymity combined with city-led control over data. In this scenario activist pressure makes the city take leadership in ensuring privacy, creating its own urban apps and minimising the extent to which its partners can share data. The tradeoff is that people must be more involved and active to get the services they need, and that criminal networks take advantage of the possibility of anonymity to flourish.

Scenario three is **rampant profiling**: high traceability combined with private-sector control over data. City data becomes a commodity in the global data market, and mainly benefits the firms that collected it. Firms are incentivised to sell data to the highest bidder, and it becomes less likely that certain groups will receive equal treatment in both commercial and citizenship operations. The final scenario is **anonymity at a price**, where a market for privacy emerges and the ability to keep one's details and activities private through encryption sells at a high price. The rich can pay to be 'greenlisted' for various forms of security, but ordinary people are tracked in ever-more detailed ways as firms try to create an incentive to pay the high price of opting out.

Findings: people, data and personal data

Most of our respondents were highly aware of the data they had volunteered – their address, workplace and information on their status and activities – but less clear about other modes of tracking. The teenagers

interviewed were the least aware, but balanced this with the highest awareness of any group about the data market and what this meant for anything they posted online. Researchers interviewed for the project outlined a long list of sources of data that already make it possible to comprehensively research movement in urban space including social media, travelcards, and wifi signals from phones.

People in the focus groups felt relatively trusting of the city with regard to its processing of the personal data that is collected when a new resident registers, one saying that 'It feels like there is a higher safety. It seems like for the fact that you know who lives where is sort of more under control and I like it more' (immigrant group). The native Dutch participants, however, felt that points of human contact with the city's data collection systems were increasingly fewer, and that they were becoming personally invisible while their data became more accessible: 'I think we are virtually invisible, at least on a digital level. I have very few touch points with the city of Amsterdam. Nearly everything I need to do with the city is … completely automated.' (technology developers group). People did feel a sense of personal contact when they engaged with the city authorities via social media, however, because 'de gemeente is meer bereikbaar op Twitter want het is publiek, iedereen ziet het'. [The municipality is more reachable through Twitter because it's public, everyone sees it.']

There was a high level of trust in every group about the city's ability to keep volunteered key personal data private, but there was much less trust regarding the new ways in which data was being produced, and the linking and merging of databases that resulted from collaborations with private partners. Interviewees frequently said that they would like to keep these kinds of data more private, or have a better idea of how they were being used, but that they felt the integration of databases would make that impossible. They had a particularly ambivalent relationship with CCTV, as shown by the contrasting responses of two immigrant focus group participants. The first said, 'I prefer to park my bike in a place with a camera than without it', and the second, 'I wouldn't like to have a camera watching me.'

The merits and perils of data for security

Most participants felt ambivalent about the automated collection of data in public space – they understood that the new data technologies could provide for monitoring and surveillance that could potentially make them and their property safer, but at the same time did not feel informed as to what was being used, how it was governed or what their role in giving permission was. In the research as a whole we found a general public doubt over what degree of public safety risk can, or should, trigger data sharing across categories and institutions. People distinguished between data that could also identify them (even if not by name or other details) and data that could not: 'If I become really personally recognisable by this facial recognition stuff, then I start to wonder what they are going to do with this. Then I would feel uncomfortable.' (technology developers group). People also made this distinction in terms of direct city services: 'I don't really want everybody to know where my car is parked. But I do want to know where there is a free spot. So I don't mind this data being used for that general benefit but I don't want it to be personal.' (technology developers group).

People also identified a problem with not understanding who was in charge of remotely collected data such as CCTV or sensor data from city infrastructure, or how it might be used and shared. They felt that it was very hard to be anonymous in a city digitally oriented toward public safety, or to choose not to participate in the city's digital life. One smartphone non-user said, 'The system that extracts and makes information has to serve those people who have nothing to hide... But some people don't want the government to know where they are.'

The risks of losing data context

One of the problems big data presents for governance is that it is stored and handled across different databases. This makes privacy self-management[1] very difficult. Our respondents felt that data use and sharing had become untrackable and hard to audit, making them highly reliant on authorities for preserving their digital privacy. Meanwhile, there is not one city administration but a collection of departments and groups, handling registration, infrastructure, traffic, building permissions, and relating to more departments such as law enforcement and the tax authorities, and data often crosses boundaries to be used in new ways. The data managed by the city may also be brought together with data from the private sector, because unless a city has internal, centralised capacity to do the data science and statistics necessary to use all the data that becomes available (as for example New York's mayor's office has established),[2] it will have to establish partnerships with commercial firms to access and analyse data, meaning that there is seldom a single actor in charge of a particular research or application using digital data.

Amongst our interviewees it was common to have experienced their data flowing across categories. Some had already experienced serious problems with public-private data flows, such as being denied a mortgage due to a particular health condition that they had only disclosed to a healthcare institution. In the case of those working in professions that were the focus of law enforcement and public safety, such as sex workers, it was common for data to flow across institutions in problematic ways, so that business registration was made public by the Chamber of Commerce (KvK) and resulted in their being denied housing, or was accessed by the highway police who then stopped them for search and questioning. This mixing of public and private space through data and registration happened because of measures designed to fight trafficking and safeguard public health, but interviewees reported that it was common for their homes to be searched by police as if they were brothels.

Thinking of the future, this kind of cross-category flow is not just problematic for sex workers, but for any workers who use public and private space in ways that have implications for regulation, such as commercial traders, taxi drivers, police, performers, builders and many others. It also shows how profiling and monitoring performed for purposes of care (combating forced sex work) can easily be experienced as control. These more regulation-sensitive jobs will only grow in the future as the labour market continues to push people toward becoming individual contractors. In this category, being flagged in a government database as risky – or, equally, at risk – leads to more prolific data-sharing across governmental departments and a greater likelihood of intervention by authorities in one's space.[3] The newer technologies that capture people's electronic signals and monitor their presence (such as smart lampposts) can also become part of systems that predict risk – who is behaving in an unusual way, who is loitering, who is gathering together in a way that might indicate unrest?

Accountability of data-organising agencies

Data infrastructures are governed both by those who set them up and use them, and by whoever is in charge of regulating them in terms of security and privacy. In Amsterdam this involves three levels of authority departmental-level permissions for collecting and using data, the city's independent commission on privacy and personal data, the Commissie Persoonsgegevens Amsterdam (the Amsterdam Commission for Personal Data, or CPA), and the national data protection authority, the CBP which generally deals with national-level issues of data misuse. In general, data protection law does not cover anonymised or de-identified data of the

[1] Solove, D. J. (2013) Privacy Self-Management and the Consent Dilemma'. *Harvard Law Review*, 126, 1880.
[2] The NYC Mayor's Office of Data Analytics (MODA): http://www1.nyc.gov/site/analytics/index.page
[3]See Keymolen, E., & Broeders, D. (2011). Innocence lost: Care and control in Dutch digital youth care. *British Journal of Social Work*, available at: https://www.researchgate.net/profile/Dennis_Broeders/publication/277524746_Innocence_Lost_Care_and_Control_in_Dutch_Digital_Youth_Care/links/55b74f7508ae092e96570f0e.pdf

kind frequently used in smart city research-and-development projects – although, importantly, location data is becoming a grey area for EU data protection regulation now that it is recognised that one's location can be both identifying and can convey details about a person's activities and personal characteristics.

Today's digital data collection and use practices present a challenge to the CPA because it does not have the capacity to proactively identify uses of data that may be problematic, but is reliant on complaints from the public or from officials. This makes it difficult for the Commission to have a preemptive function in guarding against potential harm, and also to operate on the scale on which data is now being collected and used. There is also a problem of awareness: not one person we spoke to for this research had heard of the Commissie Persoonsgegevens Amsterdam. There are also structural (and spatial) obstacles, namely that if monitoring and data collection take place in privately owned space, such as an arena, stadium or mall, only the organisers' permission is necessary. New practices are starting to emerge, however, such as the posting of information about monitoring of public space on billboards during the 2015 SAIL event.

Citizen involvement in the smart city

The language of the smart city is inclusive and collaborative: citizens are invited to take part as makers, and as active contributors of ideas and information. However, it is difficult to overcome the bias toward younger, more educated, higher-income, native-born, more technologically aware people. The people we interviewed were largely not from these demographics. They felt part of a larger datasphere whose boundaries were uncertain, but did not feel included in the planning or execution of smart city projects and research in general. They were unaware of who was organising smart city projects in Amsterdam, how they might have input, or even how to find out about what was happening. Even the highly-connected people, including technology developers, that we spoke to were sceptical about the link between digital urban smartness and participation by ordinary citizens. The city has made an effort to involve those who are interested, using a website[4] to promote a large number of public-facing projects. Yet apart from a couple of respondents who were professionally connected to smart city research or projects, the people we spoke to were unaware of them.

Our interviewees generally voiced a desire to be able to resist and exit what they saw as 'the system' of data collection and use. There were several, particularly amongst the student interviewees, who were enthusiastic about the possibilities of the smart city and digital data collection and use, but they were in a small minority. Many saw the smart city as a neoliberal, modernist project – probably because of the central role of private firms – and felt that their role as citizens was getting lost amidst a rising tide of digitisation. There is also the possibility, however, of a virtuous circle where people participate voluntarily in the datasphere in order to increase the accuracy of big datasets. One researcher interviewed said, 'I don't know if I'm scared of city planning in a data poor environment or in a data rich environment... Probably the poorer environment is again more problematic, because yeah, it's less accurate… if [people] give away their data, [at least they] know that companies and municipalities etcetera have the correct data and they're not on a false positive list of some kind.' In general the more technically involved people we interviewed were more willing to trade data on their activities, location and movement for innovations that would increase the ease with which they could move through the city, but also, by extension, participate in its life.

Conclusions: how can the city regulate the geo-information datasphere?

These are our main conclusions:

Increasing democracy

[4] http://amsterdamsmartcity.com/projects?lang=nl

- **There is an emerging democratic deficit with regard to the way data is collected and** used, and this is likely to become more of an obstacle for authorities. There are examples of census and data collection boycotts by the public that show that if public insecurity and uncertainty about data collection and management are allowed to grow, significance resistance and disruption will result.
- Our findings suggest that **creating ways for the public to connect and feed back** to public authorities may act as a check and balance on data maximisation, and may also create trust.
- **Data is essential to democratic representation**: the social contract entails people making themselves governable in return for good governance, and visibility is a necessary component of governance. But **increased visibility must be accompanied by increased trust,** something that is not currently happening. The more visible those governing data are, the more trust becomes possible.

Designing the role of the city

- Even though it only controls a small portion of the data that circulates in the massive global datasphere, **the city has great power to establish and enforce good practices to do with data**. This can be done through
    - choosing not to become a data broker between citizens and private sector contractors, but instead creating apps and services that route data only into city databases;
    - procurement, education and stimulating the regional economy;
    - managing partnerships with the private sector so that the city gets full access to data stemming from the provision of public services or collected in public space;
    - creating public consultations around the development of systems and infrastructures;
    - promoting public discussions about consent, privacy and autonomy with regard to digital data;
- It is very important to **ensure that data travels in contextual channels** and is not diverted into different purposes than those for which it was collected. This is particularly important for maintaining an ethical position with regard to the data of vulnerable or marginalised groups. This may require creating different data collection and management practices for some groups.
- **Cities that have data science capacity are better positioned to answer public concerns about data**, and to ensure data is used most efficiently and ethically.

Based on these conclusions, we suggest several paths forward.

First, **building data science capacity** within the city administration as well as partnerships with innovators outside it. Second, **engaging with political debates** about issues such as profiling, what constitutes emergency access to data, and – perhaps most importantly – how to include currently marginalised groups in the discussion about data governance. Third, **strengthening intermediary institutions** such as the Commissie Persoonsgegevens Amsterdam but also civil society organisations such as the Waag Society, Bits of Freedom and activists for participatory data such as Stichting GR1P.[5] Finally, **the rules may also need to change:** permissions regarding data reuse should be able to cross between public and privately-managed space; the city will need to **think beyond legal compliance** to consider the implications of collecting data on groups or using de-identified data, and last, **rethinking the relationship between privacy and space** so that the data production/use paradigm can take into account the emotional aspects and lived experience of privacy.

Smart city information infrastructures are in a state of emergence: it is up to those in charge to ensure that checks and balances evolve in parallel with them if the city of the future is to be not only efficient and safe, but also human and liveable. Only city governments themselves can determine whether people in the smart city will be customers, users or citizens.

[5] http://gr1p.org/

1. Introduction

1.1. Study background and aims

This report is the culmination of a year-long project conducted during 2015-16 that researched **how citizens in Amsterdam are becoming producers of digital data through their use of technology**, and the ways in which that data is becoming – or will likely become in the future – part of the way the city is governed. Our focus is primarily on spatial data, which we define as including any data that indicates a person's location or movements. This project started with a focus on geo-information (digital data with some kind of spatial tag or signal attached to it) and the infrastructures for processing it. But we have found that digital and physical spaces intermingle, and it is not possible to speak of the databases and channels for geodata without also speaking about the humans that populate the landscape, interact with it, and engage with the authorities that govern it. Therefore this report will cover issues relating to the way data is produced, managed and used by citizens, the city and the private sector, and also how people subjectively feel about those processes.

Today, most of us produce spatial data with everything we do. We get up in the morning and use a mobile phone that is constantly emitting spatial information to check our email, the news, and social media. We travel to work using an electronic travelcard or in a car with various GPS and digital systems. We walk down streets where signals from our phones and other devices are captured and read by wifi beacons, and our images by cctv. We use apps that emit details of our location, we tweet, we tag, we check in. We make phone calls through particular antennas set up by our mobile phone providers. We interact with the city digitally by paying our taxes, living in our houses, using city services and offering feedback to the authorities. All day, spatial signatures are embedded in the technologies we use, emitted as we communicate and move around, and signalled by most of our activities. This means that the picture that builds up about us in the course of every day is behavioural, but also spatial in ways that are often opaque to us. But increasingly, it is the spatial aspect of our data that tells the most detailed story about us.

Liesbet Van Zoonen has observed that city governments today are faced with a **super-wicked problem** of data governance.[6] Van Zoonen, and Levin et al. define a super-wicked problem as one where there is a perceived urgency to solving the problem, where those who cause the problem also seek to provide a solution, where the central authority needed to address it is weak or non-existent, and where, partly as a result, policy responses discount the future irrationally. Given the complex nature of the data governance challenge, then, how are authorities to determine what elements of public life are private processes, and how can they be balanced with the benefits of sharing to create public goods, such as health and education?[7]

We posit that in order to do so, it is first necessary to conceptualise what privacy means with regard to big data – something that both national and international authorities have not so far been able to do. Yet channelling and managing 'big' digital data is going to become a particular problem for those managing urban environments. The spatial data we produce has, until recently, been used only incidentally by municipalities.

---

[6] Van Zoonen, L. (2015). Big, Open and Linked Data (BOLD) challenges for urban governance. Paper presented at the Data Power Conference, University of Sheffield, June 22-23, 2015. Following Levin, K., Cashore, Be., Bernstein, S. & Auld, G. (2012). Overcoming the tragedy of super wicked problems: constraining our future selves to ameliorate global climate change. Policy Sciences 45 (2): 123–152.

[7] Here we quote Liesbet van Zoonen's paper, *Big, open and linked data challenges for urban governance*, available at: https://www.researchgate.net/publication/279190937_BIG_OPEN_AND_LINKED_DATA_CHALLENGES_FOR_URBAN_GOVERNANCE

But today in the fields of research and government, such signals are becoming increasingly useful as ways to track and monitor what is happening in urban space. **Cities themselves are creating data on us, sharing it, and using it to provide us with services, track us and communicate with us**. Sensors such as automatic number-plate recognition (ANPR) track millions of cars daily, and we are monitored as we use ATMs, credit cards, work in smart buildings and live in smart homes. Many cities worldwide are adopting data science labs as key tools of urban governance: they integrate and analyse the data that emerges from all the city's systems with the aim of producing insights that will make the city more efficient, safer and better to live in.[8] At the same time, city infrastructure such as smart street lighting, CCTV and even waste bins[9] are equipped with sensors that collect identifying signals from phones to track people as they move through urban space.

In the future, it is likely that these data generated by city infrastructure and registration systems will become merged and linked with data generated purposely by city residents. People's social media postings, data from self-tracking devices and smart homes, maps generated by crowdsourcing, and feedback of all kinds will be merged to create a more complete picture of the city's functions and dynamics. One of the researchers interviewed as an expert for this project noted that ten years ago, it was estimated (by the Dutch CBP) that every Dutch person's details were in somewhere between 500 and 3,500 databases. Ten years later with the era of big data firmly established, both the number of databases and the technical capacity to link them together to produce new insights and profiles have increased exponentially.[10]

At the same time, there are many who are left out and disempowered by the increasingly datafied city. Recent research from the Dutch bureau of statistics shows that 1.2 million people 12 or older in the Netherlands are non-users of the internet. A majority of them were older people, with women and the less-educated most represented. If we are supposed to engage with the smart city through our use of technology, this currently makes it impossible for eight per cent of the country to participate.[11]

From research on the development of information systems and infrastructure, including geo-information and spatial data infrastructure, we have learned that it is difficult to design a fully functioning system from scratch, but rather that infrastructure emerges through an interplay between technical and human agents. This process often has unpredictable paths and results in political choices becoming hidden in the system's nature and logic, so that the political impacts on society are difficult to trace[12]. Information infrastructures have long left the boundaries of formally circumscribed organizations, such as one company, one government department, or even one country (as in the case of national information systems). As technological development takes on increasingly diverse forms and increasing speed, for instance in smart cities, insights into the governance of such systems and how they in turn govern society are lagging in an alarming manner. This project seeks to contribute to understanding the development of geo-information infrastructure in Amsterdam with specific focus on the citizen's perspective and the implications these bear for political and regulatory processes and dialogue.

We approached this project with the view that people's increasing awareness that they are producing digital data will change citizenship practices – they will relate differently to those who have access to their data, and will start to think about how they want their data channelled and managed. The authorities' linking and

[8] Examples include New York (http://www1.nyc.gov/site/analytics/index.page) and Dublin (http://www.dublindashboard.ie/pages/index)
[9] http://www.wired.co.uk/news/archive/2013-08/09/recycling-bins-are-watching-you
[10] Sargasso found in 2012 that the number was at least 5,000 (http://sargasso.nl/meer-dan-5000-databases-met-persoonsgegevens-bij-overheid/)
[11] https://www.cbs.nl/nl-nl/nieuws/2016/22/acht-procent-van-de-nederlanders-nooit-op-internet
[12] For example, Avgerou, C., & McGrath, K. (2007). Power, rationality, and the art of living through socio-technical change. MIS Quarterly, 31(2), 295-315, andStar, S. L. & K. Ruhleder (1996). Steps toward an ecology of infrastructure design and access for large information spaces. Information Systems Journal, 7(1), 111-134.

merging of datasets (such as tax data with location records, travel data with social media postings) will be debated. So will the extent to which people's 'public' data (such as social media, check-ins, and other postings on public platforms) is truly volunteered, and may be used in any context, versus being observed or derived data and therefore subject to permissions.[13] The question of commercialisation of data will also become important as the city interacts with commercial partners who are interested in using and selling data more broadly, and so will the ways in which existing rules such as the Fair Information Practice Principles (FIPPS)[14] can be applied to data flows in urban governance.

Box 1: the 'Rode Loper' crowd observation project

In Amsterdam, an example of the kind of monitoring that is possible in city space is the Rode Loper project, which took place in 2014 as a collaboration between the City of Amsterdam, researchers from TU Delft and private sector consultants Dat.Mobility. The project constituted cutting-edge research, and involved monitoring pedestrian and cycle traffic in the 'Rode Loper' zone between Central Station and Dam Square. On the busy shopping street Kalverstraat, during two months of 2014, all passers-by were tracked in several ways: first, by CCTV with facial recognition technology; also by wifi beacons which picked up device-identifying signals from mobile phones, and finally by de-identified mobile phone data purchased through an intermediary firm, Mezuro, which transacts mobile phone information from various operators. All this created a detailed picture of the traffic patterns in the Kalverstraat over the two-month period. The research was not communicated to the people passing through the area, so that those being monitored were unaware of the process.

The Rode Loper project is a good example with which to begin our report, because it outlines several of the problems inherent in the evolving capacity to track, monitor and record our movements and activities. The technology used to recognise people in a crowd is no longer esoteric or restricted to the security services, but is becoming accessible to everyone.[15] Under what circumstances should the city institute a system to track and monitor people's movements and activities? What kind of data should be collected and who should have access? Should it be available in real time, shared, repurposed, stored for the longer term? How should the people being monitored be made aware of the system, and should they have power over what is done with data that reflects them? These questions become especially pertinent with projects such as this example, where the data in question is not clearly defined as personal – i.e. it does not clearly relate to individual identities. Does it matter that the data is de-identified, and is tracking a presence the same as tracking an individual? Does it make a different that the tracking is for academic research purposes and that there is no commercial aim? How should the city weigh the public safety and innovation aspects of the project against those of preserving people's privacy? And perhaps most importantly, who should decide on all these questions?

This report seeks to answer two main questions. First, **what organisational and governance structures are necessary for a sound and innovative spatial data infrastructure in Amsterdam?** And second, **how can the city address the needs, privacy rights and responsibilities of citizens who create and use spatial data?** These

[13] These definitions come from Hildebrandt, M. (2013). Slaves to Big Data. Or Are We?. *Idp. Revista De Internet, Derecho y Política* 16.
[14] https://en.wikipedia.org/wiki/FTC_Fair_Information_Practice
[15] https://www.theguardian.com/technology/2016/may/17/findface-face-recognition-app-end-public-anonymity-vkontakte

questions have no simple answers, but they led us to seek to include in this project groups and experts who were outside the mainstream or actually marginalised in various ways within the city. We know from information infrastructure theory that such infrastructure always excludes some groups of people[16]. The smoother and more seamlessly the infrastructure runs internally and the more tightly integrated it is, the more strongly it tends to exclude some groups of people and technical elements. We address the idea of a 'sound' data infrastructure as one that is fit for a future where people make different choices about engaging with digital technologies, in a world where the city is becoming more programmable and digitally integrated as technology continues to advance.[17] What kind of governance of digital data creates an equal playing field for the elderly, the young, the vulnerable? For non-users of smart technologies, non-citizens, speakers of other languages?

With the idea that infrastructure emerges rather than being designed, and that city data infrastructures are currently emerging, we want to lay the ground for developers, city authorities, firms and researchers to ask the kinds of questions that can provide a basis for envisioning an equitable governance structure for spatial data in Amsterdam that will be adaptable to future developments and responsive to residents' needs and opinions. We also aim to pose questions about transparency, accountability and use of the data that will make it possible for citizens and contributors of information, as well as municipal authorities, to be beneficiaries of that infrastructure.

1.2. Methodology

This research was conducted through interviews and observation in Amsterdam over the course of 2015 (for a more detailed explanation, see Annex). Over the course of 2014-15 we also participated in a range of events and discussions to do with smart city technologies, charting the actors engaging in the field from different sectors, and the debates emerging. The project was conducted partly in Dutch but mostly in English: the expert interviews were conducted in English, but the events we attended and the focus group discussions were often in a mixture of Dutch and English, and about 20 per cent of our respondents spoke only Dutch during the research. We began with a series of 20 interviews with experts in data and urban governance, followed by a scenario-building exercise in which we used that information to think about what data use scenarios might look like in the future. We then conducted a series of eight focus groups, with 6-10 people in each. These focused on issues and types of citizen, with a particular emphasis on those who might be disadvantaged by, or were likely to be particularly sensitive to, an increase in the city's reliance on digital data as a way to relate to and work with residents. These groups were 1) people at higher risk of being profiled; 2) non-users of smart technologies, 3) sex workers, 4) non-EU immigrants, 5) EU immigrants, 6) freelancers (ZZP'ers), 7) technology developers and 8) schoolchildren. In the focus group process our role was to make people aware of the current and future possibilities of spatial data, and then see how they responded. The focus groups were convened using contacts from the entire research team (10 people in all were involved in the project) and were chosen based on the research team's background research and discussions.

In convening these focus groups, we aimed to determine which groups were currently missing from, or marginalised by, current discussions and practices of smart city development, and also those whose lives might be changed most by an increase in urban datafication. Our discussions highlighted several groups: non-natives; ethnic or religious minorities; children and the elderly; those who opted out of using the technologies currently seen as necessary for citizen involvement in the smart city (i.e. smartphones); those who operate in highly

[16] Aanestad, M., E. Monteiro, & P. Nielsen (2007). Information Infrastructures and Public Goods: Analytical and Practical Implications for SDI. Information Technology for Development, 13(1), 7-25. and Star, S. L. (1999). The Ethnography of Infrastructure, American Behavioral Scientist, 43 (3):377-391.
[17] Kitchin, R., & Dodge, M. (2011). *Code/space: Software and everyday life*. MIT Press.

regulated professions, and freelancers who are responsible for their own working environment. In this report we use the words 'citizen' and 'resident' interchangeably to signify people living and working in Amsterdam who are invested in the city's development over the long term.

The structure of this report is as follows. First, we outline the types of data we focused on, and the kinds of question they raise. Next, we explain the scenarios we came up with for future urban uses of smart technology and their implications for people living in the city. We then explore our findings from our expert interviews and focus groups in terms of the four main issues that emerged: 1) people's increasing visibility through new digital technologies; 2) arguments for security, efficiency and how these play out in terms of data collection and use; 3) who is responsible for decisionmaking about the city's use of digital data, and 4) the involvement of city-dwellers in the innovation of these technologies and related decisions. We conclude with an overview of our findings and recommendations for city governments interested in exploring these questions further.

2. Framing the problem

2.1. Initial observations: data and the city

From our expert interviews, we sought to understand what can currently be known about us from our digital lives, and how this is likely to develop in the coming decade. The interviews focused both on the data that is currently available to government, and the implications of future merging and linking of different types of data to those established databases. Together, the interviews outlined a future where data of all kinds is increasingly traded and merged to provide a multifaceted picture of how the city is operating. The experts we consulted often related the availability of data to the ability to see the city's dynamics, rather than those of the individuals in the city. In the datafied city we discussed with these experts, we frequently found that individuals became objectified and were seen as incidental to data flows – rather than living parts of the city's operations and dynamics, they easily became problems to be solved (public safety or public health risks), or groups to be influenced and controlled – users of the city, rather than its living infrastructure. This was particularly true when we looked at data flows from a spatial perspective – geodata is used for policy analysis, creating interventions and understanding the larger scale flows and moods of the city, making it easy to lose the sense that those flows are made up of individuals with agency and preferences.

At the moment, mobile phones are probably the richest source of information on our movements and activities. To outline how much information on us is available through our use of our phones,

Table 1 below shows the kinds of data that can be requested by any app running on the Android operating system[18] and Figure 1 shows the way that a GPS sensor in a phone can allow the tracking of individual movements. For figure 1, the picture is similar between Apple and Android phones, but for table 1 the information accessible to app developers will differ (and is often less extensive) for Apple phones.

Table 1. Data collection by Android apps

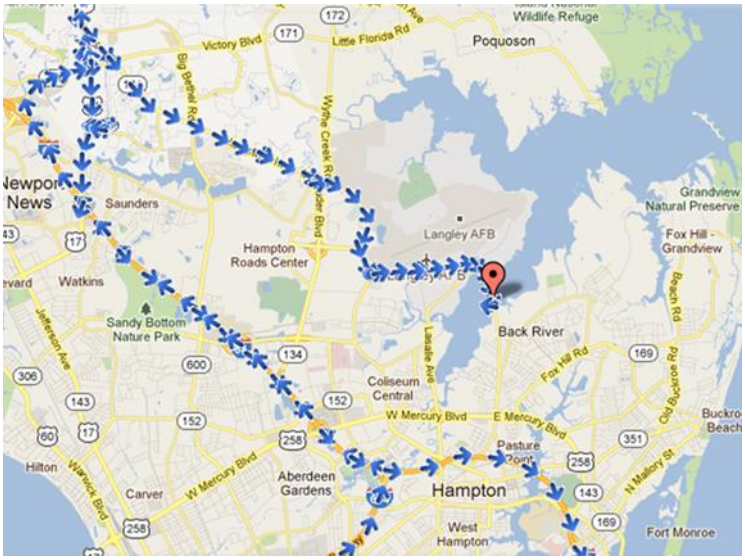
Data type	Data permissions sought
Accounts log	Email log
App Activity	Name, package name, process number of activity, processed id

[18] This list is taken from Hein (2014): http://www.cultofmac.com/304401/ubers-android-app-literally-malware/

App Data Usage	Cache size, code size, data size, name, package name
App Install	Installed at, name, package name, unknown sources enabled, version code, version name
Battery	Health, level, plugged, present, scale, status, technology, temperature, voltage
Device Info	Board, brand, build version, cell number, device, device type, display, fingerprint, IP (internet provider), MAC address (device identifier code), manufacturer, model, OS platform, product, SDK (software development kit) code, total disk space, unknown sources enabled
GPS	Accuracy, altitude, latitude, longitude, provider, speed
MMS	From number, MMS at, MMS type, service number, to number
NetData	Bytes received, bytes sent, connection type, interface type
PhoneCall	Call duration, called at, from number, phone call type, to number
SMS	From number, service number, SMS at, SMS type, to number
Telephony Information	Cell tower ID, cell tower latitude, cell tower longitude, IMEI (International Mobile Equipment Identity), ISO country code, local area code, MEID (Mobile Equipment Identifier), mobile country code, mobile network code, network name, network type, phone type, SIM serial number, SIM state, subscriber ID
WifiConnection	device identifier codes: BSSID, IP, linkspeed, MAC address, network ID, RSSI, SSID
WifiNeighbors	device identifier codes: BSSID, capabilities, frequency, level, SSID
Root Check	Root status code, root status reason code, root version, sig file version
Malware Info	Algorithm confidence, app list, found malware, malware SDK version (software development kit), package list, reason code, service list, sigfile version (signature for communications sent by user)

Figure 1 shows the way that a detailed trajectory of a person's movements can be tracked through the GPS sensor of their mobile phone – in this case by an employer using an app loaded onto their employee's phone. Employee tracking is more common in the US than the EU, but is gaining increasing supporters amongst European employers at least partly because of the current heightened security risks of urban European life due to terrorism and violence. Every mobile phone can be tracked spatially regardless of whether it has a tracking app loaded, either through its regular checks with the network through antennae in a particular area ('coarse location' in technical terms) or through its communication with wifi networks which provide a much more specific location tracking ability.

Figure 1. Daily trajectory of a mobile phone, measured by GPS sensor



Source: Employeetracker, Frontrangepremiergroup.com

The examples above are of data collected, managed and used by the private sector, which currently controls most of the location data available (by volume, since massive amounts stem from people's use of digital devices). However, in the future it is likely that data generated by people's devices will become more available to the public sector, that there will be new norms, rules and alliances for data sharing, and that data will increasingly flow between sectors. There is currently an important distinction between data that shows people's identity and data that is de-identified (either by replacing names and phone numbers with numbers or other markers, or by removing those identifying features from the dataset entirely). Despite anonymisation, however, it is still possible to tell a lot about a person from their phone location data – where they sleep at night, where they spend most of their time during the day, the kind of places they visit. Currently de-identified data is traded by intermediary firms, and are often used by cities to analyse people's movements and activities. In the future we can expect to see more merging of private with public data, and collaborations between the city and private sector interests. Another participant in this process is likely to be academia, since academic researchers frequently, and increasingly so, form a bridge between public and private sector with regard to digital data analytics.

Table 2 below shows the types of data currently available to and from city authorities. It is not an exhaustive list, but covers the main categories of information that are currently being shared or negotiated between government, academia and the private sector.

Table 2. Types of data available to/in cities

Type of data	**Who may have access**	**Who can gain access if authorised**
Mobile phone records	When identifiable, provider only. When de-identified, often available commercially and for academic research	Police, security and emergency services
Social media postings	Public	Public
Volunteered geographic information (e.g. crowd-mapping)	Public	Public

Identity and address information (GBA)	Government, landlords, commercial firms (e.g. newspapers for deliveries)	Police and security services
CCTV (Closed circuit television)	Owner of service (private or public), research institutions	Police and security services
ANPR (Automated Number-Plate Recognition) – traffic flow information	Government, research institutions	Police and security services
Car GPS data	Owner of service (private firm), also traded in de-identified form to government and other firms	Police and security services
Tax records	Government institutions, anti-fraud	Police and security services
Bank records	Banks, tax authority, anti-fraud	Police and security services
Travelcard information	Firm owning cards; data also shared under contract with transport authorities, and in de-identified form with government, firms, interest groups, research institutions	Police and security services
Wifi data	Commercial companies, government and academic researchers	Police and security services
Healthcare data[19]	Healthcare providers, insurance companies, de-identified data used by health authorities and research institutions	Police and security services
Pensions and benefits	Government, pension funds, employers, anti-fraud	Police and security services
Driver licensing and vehicle information	Provider, government departments	Police and security services
Utility service records (water, electricity, gas)	Utility providers, research institutions, police, tax authorities, anti-fraud	Police and security services
Education records	Education providers, city government, school inspectors, research institutions	
Criminal records	Police, screening organisation within government	

Source: authors & Prof. Dennis Broeders, EUR.

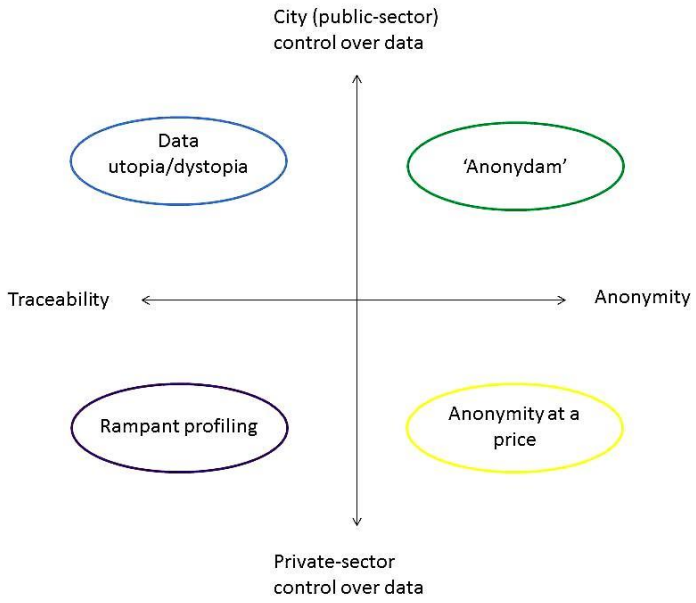
The table shows some of the flows of data taking place, but cannot cover the complex structures of databases and agreements that govern the availability of different types of data. For a more complete discussion of the Dutch case, see the WRR's i-Government report: in Dutch at http://www.wrr.nl/fileadmin/nl/publicaties/PDF-Rapporten/I_Overheid.pdf and in English at http://www.wrr.nl/fileadmin/en/publicaties/PDF-samenvattingen/iGovernment.pdf.

[19] Health data is increasingly generated in parallel by individuals, using 'quantified self' technologies such as fitbits and health tracking services. This data is sometimes, but not always, traded on the open market depending on user permissions.

2.2. Scenarios for digital data in the future city

Based on the information available from both background research and our expert interviews, the research team built several scenarios for the way data may be used in the city of the future. These scenarios are extremes: they are drawn from a conceptualisation of two key axes along which data use and permissions may develop (see figure 2). The scenarios then fed into the initial information and questions posed to the focus groups, and served as a guide for the discussions. They were particularly important in orienting the discussion toward a broad perspective on urban data uses and the kinds of data that might be spatial, since the focus group participants naturally tended to refer to technologies already in play and that they used on an individual level every day. These 'bigger picture' scenarios served two purposes. First, constructing these scenarios in a workshop was the first round of analysis of expert interview data. Second, the resulting scenarios were subsequently used by the researchers to support participants to imagine possible technological futures and the kinds of urban code/space they would prefer to live in, or were nervous about seeing evolve.

Figure 2. Urban data scenarios



1. Data utopia/dystopia: high traceability and city-led data control

This scenario imagines that people have become highly visible through data that is openly available to municipal authorities. The city uses data on everything from people's movements to their electricity usage at home to chart their lives and activities, and monitors both public and private spaces in real time. Authorities are able to profile people in great detail and to target policies and services on a neighbourhood or even household level. A decreased feeling of individuality in people's urban lives mean citizens feel more like members of particular groups that are addressed by the city in specific ways.

This leads to real progress in ensuring equal representation within the city: the needs of the marginalised are clearly expressed to the authorities and people perceive the city as treating them more equally. Public safety also benefits, as the city has a real-time feed of data on everything from traffic problems to crime. The city also becomes more environmentally friendly and sustainable because people, and their fellow citizens, can

immediately tell when they are acting wastefully or over-using resources and can thus act to regulate their own and others' behaviour.

The scenario also has negative aspects. Social engineering is made easier and abuse more possible. People often feel exposed, and may find some data sharing abuses their privacy. Social tensions build around the city's monitoring practices, setting citizens against the authorities and necessitating new activist organisations to fight for people's rights not to be surveilled. Constant monitoring also leads people to behave as if they were being watched all the time, censoring their own behaviour because they feel the eyes of the authorities on them in public and private space. People feel less like individuals and more like members of the herd, and an increased sense of order ends up decreasing creativity overall.

2. 'Anonydam': anonymity and city-led data control

In this scenario the state allows a lot of data to be collected and shared about people's lives and activities, but due to activist pressure through social networks the city government takes a leadership role in ensuring the privacy of its citizens, exerting its power as much as possible to give people greater anonymity. It does so through its role in governing the 'smart city' applications which feed back data in a separate loop to the city rather than the national government.

The city processes all the data it receives in ways that preserve anonymity at the stages of collecting the data, analysing it, and reporting and acting on it. The city therefore gives up the opportunity to know about people's movements, activities and behaviour in greater detail, and also gives up some of the opportunity to intervene in their lives.

The result is that people are not easily categorised and known by the authorities. This means citizens must build local social networks in order to relate to each other and make decisions about their neighbourhoods and city policy. This increases democratic engagement and confidence, as people become more involved and activist. It also safeguards privacy, and sets an example for the national government at a time when detailed data is becoming ever more available. People interact with their government in ways that they choose, and the city government's accountability to its citizens increases.

The city's protection of anonymity also has some negative effects: criminal networks are able to flourish because people are not tracked or monitored. (This scenario supposes that the current discourse about the importance of anti-terrorism has waned and public security is no longer the primary justification for collecting data.) The city's privacy policy limits firms' ability to market their products to people, since they are less able to categorise people for advertising purposes. It also places some restrictions on data-driven innovation such as apps for smartphones and online services.

3. Rampant profiling: traceability and private-sector control

In this scenario, data is primarily collected and controlled by private firms working for the city, but the city does not have access to most of the data they collect. The data becomes a commodity in the global data market, and mainly benefits the firms that collected it. In the city, all residents are profiled individually in great detail through their use of personal devices and through their presence in smart environments, where they are tracked by sensors - but the data is not used unless money can be made.

The positive side of this scenario is that the data people emit is handled within corporate database systems and is subject to data protection law, which tends to be more stringent for private-sector data controllers than for governmental authorities. The data is treated by firms as an important link with their clients, and is therefore handled with care in order not to lose business and suffer reputational damage.

The negative side of this scenario is that firms are incentivised to sell data to the highest bidder, and that data protection law still leaves many loopholes for the use of anonymised data in particular, which can still be used to profile and potentially discriminate against groups. It becomes less likely that certain groups will receive equal treatment in both commercial and citizenship operations: there are restrictions on the availability of insurance products, mortgages and loans on the basis of personal characteristics – health, household composition and personal history, and the prices charged for all kinds of products vary based on people's income and assets. Furthermore, public-sector decisionmaking becomes based on profiling done by the private sector data controllers, so that people can be blacklisted as security risks based on their social networks, movements, activities or financial transactions. The processes of blacklisting (denial of access), greenlisting (providing access) or greylisting (marking as risky)[20] are black-boxed and cities have to choose whether to use the resulting findings or not without fully understanding the processes that have created them.

4. 'Anonymity at a price': privacy under private-sector control

In this scenario, the private sector holds detailed data on people that is not available to public authorities. The firms that collect and process the data perceive a demand for anonymity and therefore create a market for it. In this market, the ability to keep one's details and activities private sells at a high price, so that the minority with the most resources can become invisible at the expense of the majority. Conversely, ordinary people are tracked in ever-more detailed ways as firms try to raise the value of privacy and create an incentive to pay the high price of opting out.

The first result of this is an increase in inequality, both practically and in perception. The rich can afford a cloak of invisibility where they are relatively unaccountable to the state or city, while the poor are tracked in detail and become more subject to control in an attempt by the authorities to govern everyone by governing those they can see. The rich can afford encryption services for all their communications and transactions, using products such as the Blackphone,[21] and the online currency Bitcoin. The rich can also pay to be 'greenlisted' for various forms of security – airports, transport, privatised city services – so that they get priority and can travel and function more easily.

Meanwhile the poor are profiled and targeted for ubiquitous direct marketing and advertising through the content of their communications and activities. The state and city have to buy people's information from the private sector, but the lack of an open data program makes them less accountable for their use of the data.

Other markets develop as a response to extreme private-sector control over data: a black market for even greater anonymity, where criminal networks provide invisibility for a high price to those who want to engage in illegal trade or activities, ensuring privacy from the firms who collect data and could sell it to law enforcement. Because data can now be sold to the state or city for population registers, taxes, and other citizenship functions, companies enforce 'real identity' policies online to ensure they are getting accurate data on people. This leads to the emergence of another market in online pseudonymity, making people less visible to tax authorities and others interested in tracing them.

2.3. Emergence of a geo-information infrastructure

The current shape of the geo-information infrastructure (the technologies and organisational channels used to collect, process and analyse information about us that relates to the way we use or occupy space) we see

[20] Broeders, D., & Hampshire, J. (2013). Dreaming of seamless borders: ICTs and the pre-emptive governance of mobility in Europe. *Journal of Ethnic and Migration Studies*, 39(8), 1201-1218.
[21] https://www.silentcircle.com/

emerging in Amsterdam is characterised by fast technological development, the multiplicity of actors and organizations not only already involved, but also continuously joining through various debates, pilot projects, and events organized around concepts such as the 'smart city'. New actor alliances are emerging at a relatively fast pace between government, private sector, and research institutions. Visible nodes and elements of this growing multi-layered network include so-called living labs and experiments with the digitalisation of urban life (see box 2). The geo-information infrastructure which we see emerging in the context of the smart city and around the digitisation of urban life differs from more traditional geographic information systems in administration and spatial data infrastructures in terms of the multiplicity of interests, where government is currently taking on only one of many roles, and in the speed at which technology-driven ideas are added and experimented with. It is a geo-information infrastructure in the era of big data.

One of the reasons that big data presents new problems in terms of governance is because it is distributed: it exists across databases, applications, owners and permissions. Another important difference between more traditional geo-information infrastructure and the infrastructure we see emerging in our study is that spatial data is just one strand within many different data flows. Geographic information systems have for a long time run on the principle where spatial data is stored in one format and attribute data (characteristics about the space) are stored in another format, then the two are linked within a geo-database (for use with mapping software) or database network. In the current geo-information landscape of the city, however, data is no longer subject to this binary division. People's location and movement are captured in many different formats and as data embedded or derived from other data flows. How one's location and movement are being mapped in the era of big data analytics becomes increasingly complex and unpredictable. The difference between spatial data, on one hand, and attribute data, on the other, is dissolving, as is the difference between digital space and non-digital. In the era of big data our digital whereabouts can tell a lot (sometimes everything) about our physical location.

Given the changes and complexity of the emerging geo-infrastructure, we used the four possible future scenarios outlined in the previous section as the basis for our focus group discussions in order to find out how city residents see - or cannot see - the development of this new type of geo-information infrastructure, its risks and potentials.

Box 2: Living labs

The idea of a city 'living lab' initially implied real-time experiments that often involved human participants in urban environments. The term has evolved to include various different models, but centres around the insertion of a new technology-mediated practice, digitally enabled infrastructure or network in a particular area of the city, in a context where the process can be monitored and evaluated. Crowd management pilots, smart lighting experiments, energy-efficiency projects, experiments to influence or change people's behaviour and neighbourhood-focused technology-driven projects aiming for social change all fall under the definition of living labs.

Living labs may involve different degrees of citizen participation and awareness. People may participate by downloading apps that track their movements or behaviour, in which case they are consenting partners in the research. They may be tracked in ways that are clear to them, for instance if a project is publicised in advance and consultations are held or feedback invited (see the case of the CPA, for instance, in section 2.2.3). A third model involves remote data collection without notifying people that they are participating in a research project or experiment. The third model is currently the norm, mainly because data protection law does not apply to data that does not make people personally identifiable, so that authorities and researchers are not obliged to make people aware if they are collecting data in public space.

3. Findings - Citizen perceptions across four themes

The next sections look at the findings from our focus groups in conjunction with material from expert interviews. We present more detailed themes relating to the problem of data governance, bearing in mind the future scenarios. We began the majority of the focus groups with an exercise where participants were invited to list anyone, commercial or public, personal or unknown to them, who might know personal details about them at that moment, and what they knew. We also asked who might know their location. We found that most were highly aware of who knew their home address, usually as part of administrative details given consciously by the interviewee. However, they were less clear about who might have their location details. This difference was particularly pronounced with the group of teenage schoolchildren: a collective exercise showed that the participants could name 12 different parties or institutions that knew their home address (including family, neighbours, employers, school, the city government and their bank), but were much hazier about how their location became known electronically. Throughout all the focus groups, there was a marked absence of awareness about the extent to which mobile phones make their users visible and trackable in various ways.

3.1. Security, efficiency and data maximisation

One of the themes that emerged from the interviews was a tension between the aims of efficiency and security, which led the city to install 'smart' datafied systems that generate and process data in new ways, and the need to preserve people's autonomy and anonymity in their daily lives. Moreover, people felt that data about them was constantly leaking from their digital lives, and that there was no ethic of minimisation in terms of data collection – everyone around them wanted all the data possible. The tension between smooth running of the city and minimising monitoring is a familiar one from surveillance theory, which posits[22] that **monitoring and surveillance can be seen as a continuum, with aims of care at one extreme and control at the other**. Projects may begin at the monitoring-for-care end of the continuum, but experience 'function creep' towards control as their other uses become clear. When faced with the idea of connected infrastructures that could track their activities and their use of resources, for example smart lampposts, smart electricity meters or wifi capturing beacons in public space, interviewees were clearly able to perceive both their potential for service delivery and for monitoring behaviour.

3.1.1. A digital observatory of city life

The problem of data maximisation grows with the variety of data available. Researchers interviewed for this project were asked to name the kinds of data they would like in order to do their ideal crowd management analysis: their response was that the ideal research tool would be a crowd management dashboard with real time information on the situation in the city, 'particularly at pinch points where you expect problems in terms of safety', with sensors to capture 'wifi and bluetooth and video, and cameras that can count, and infrared and GPS and twitter.' They would also like for each person in the city 'a trajectory with an activity chain… the activities that people take part in that cause them to move from A to B and in the end wind up in a traffic jam'. This trajectory would be made up of traditional traffic data such as roadside census data, license plate cameras to provide travel times between specific linked places, combined with data on the city infrastructure itself such as metrics from intersection controllers. The researchers also wanted 'the cooler stuff, so if you have people that drive around with a navigation device, the GPS data, and all types of event-related data on what people are twittering, and also mobile phone data as far as we can get our hands on it… And then of course public transport data, the OV Chipkaart data.'

3.1.2. Public trust in data collection and use

People in the focus groups felt relatively trusting of the city with regard to its processing of the personal data that is collected when a new resident registers. These data include name, address, occupation, immigration status and civil status. Some, particularly immigrants who had to register for the first time as adults, liked being registered and commented that the system of municipal registration felt like a safety net, and that they liked the feeling that their presence had been registered:

> 'It feels like there is a higher safety. It seems like for the fact that you know who lives where is sort of more under control and I like it more.' (immigrant focus group, 2.12.15)

Residents of Dutch origin, however, felt that the increasing digitisation of the municipal registration system meant that there was little contact on a human level about people's data:

> 'I think we are virtually invisible, at least on a digital level. I have very little touch points with the city of Amsterdam. Nearly everything I need to do with the city is either completely automated without

[22] Lyon, D. (2008). Surveillance Society. Presented at Festival del Diritto, Piacenza, Italia: September 28 2008.

> any touch points or the rare following up about your driving licence or passport. That's all.' (technology developers, 3.12.15)

This point of view is interesting because it contrasts an increasing sense of personal invisibility in digitised systems with the increasing digital visibility that comes with our emission of more types of data. In such systems, the city bureaucracy can identify us, but it cannot see who we are.

There was a high level of trust in every group about the city's ability to keep volunteered key personal data private, but there was much less trust amongst interviewees about the new ways in which data was being produced, and the analytics and merging of databases that resulted from collaborations with private partners:

> 'I think we don't know how much integration there is already taking place. So we have Park-Mobile, which is of course linked to some part of the government or an external agency that actually managed the whole parking place payment stuff. I don't know whether an extract of that is actually provided to the government as well. I don't know, but I think with all the digital services that are provided within the city it depends on how much access the municipality has to this information. (technology developers, 3.12.15)

> '[the city collects our data] to assess what services are needed – transport capacity, disease reporting – I don't feel spied on by the state. If I go to a public service place and they have my data, that's alright. I mind if companies have my data.' (non-users, 10.9.15)

Interviewees frequently said that they would like to keep these kinds of data more private, or have a better idea of how they were being used, but that they felt the integration (or possible integration) of databases made that impossible:

> OV [chip]kaart as well, it generates a lot of data on where you have been... Therefore, I have a non-personal [card] but because I pay the fee on it with my bank card probably they do know. (technology developers, 3.12.15)

Focus-group participants had a particularly uneasy relationship with CCTV. The statements of two European immigrant interviewees encapsulate this uncertainty:

> Interviewee 1: I prefer to park my bike in a place with a camera than without it.

> Interviewee 2: I wouldn't like to have a camera watching me. (Immigrants, 1.10.15)

These statements encapsulate the ambivalence many participants felt towards the collection of data in public space – they understood that the new data technologies could provide for monitoring and surveillance that could potentially make them and their property safer, but at the same time did not feel informed as to what was being used, how it was governed or what their role in giving permission was. Under these circumstances it may be considered natural that people felt intuitively hostile to the idea of surveillance while also being able to imagine that it might confer benefits.

3.1.3. What justifies using data for public safety?

The Rode Loper project (see Introduction) brought up some questions about how data about crowds, such as CCTV and facial recognition data, should be stored, managed and used. In the focus group composed of those who did not use smartphones, there was disagreement about whether, if cameras in public space captured a crime occurring, the data should then be used to investigate. In a hypothetical case where either pickpocketing

had occurred in the area being filmed, or where a woman had been beaten up in the street, in each case two participants felt it was alright to use the data and five felt that it was not. In each case, several people felt that it was too extreme a repurposing of data for too little potential return, with comments including 'the police never get pickpockets', 'there should be more prevention', and 'I have been in that situation and they didn't catch anyone'. In the research as a whole we found a general public doubt over what degree of public safety risk can, or should, trigger data sharing across categories and institutions (particularly with law enforcement), and as to what the rules were with regard to security and sharing data.

When asked how comfortable they felt being tracked for crowd management purposes, interviewees distinguished between systems that could recognise individuals and those that could not.

> '…[if] you don't have the facial recognition, it's just used to know where you are walking as an unidentified walking object, and it is used for all these benefits, and I can see actually on the app where it's busy and where not, then it benefits me without me being personally recognisable. If I become really personally recognisable by this facial recognition stuff, then I start to wonder what they are going to do with this. Then I would feel uncomfortable.' (technology developers, 3.12.15)

People also made this distinction in terms of systems that they could see and use directly, such as a hypothetical city-run app for finding parking spaces:

> 'I don't really want everybody to know where my car is parked. But I do want to know where there is a free spot. So I don't mind this data being used for that general benefit but I don't want it to be personal.' (technology developers, 3.12.15)

Public safety measures are partly performative. The performance of certain actions – taking off our shoes at the airport, producing identification on demand – is designed to give rise to a sense that something is being done to keep us safe. However, when presented with the decontextualised news that CCTV was being used for crowd control brought up the problem that they did not feel more secure just because more data was being collected. In fact, rather than remote data collection, what made interviewees feel secure and engaged was their direct contact with the city government, performed through acts of registration and moments of contact over service provision and feedback. This included complaints – people felt they did not know who was in charge of remotely collected data such as CCTV or sensor data from city infrastructure, but were willing to engage with the city authorities via social media because 'de gemeente is meer bereikbaar op Twitter want het is publiek, iedereen ziet het'. [The municipality is more reachable through Twitter because it's public, everyone sees.']

Several participants, including but not limited to those who did not use smartphones, felt that it was important to preserve the city as a space where people could be anonymous, or could choose not to connect to the city's digital life. One participant said,

> 'The system that extracts and makes information has to serve those people who have nothing to hide. But what about people who have some things to hide or want to keep things personal? There is going to be a system that knows where somebody is and how many times a day he crosses the road and it's important to the traffic flow... and almost all people will be OK because they can walk safely and the flow goes very well. But some people don't want the government to know where they are.' (smartphone non-users, 10.9.15)

The question of what we may keep secret and what is public is germane to city life in particular. Urban environments have always generated debates about how people can live together without invading each other's space. The addition of electronic means of defining, coding and making space serve various interests

and informational needs at the same time adds a new layer to this debate, and challenges authorities to find new ways of delineating and informing city-dwellers about how the public space around them is defined and used.

Box 3: Smart lighting in public space

Smart lampposts now populate the area around Amsterdam's Bijlmer Arena, an area that frequently sees crowds entering and exiting sports games and events. The lampposts sense how many people are in the area and turn the light up or down accordingly. They can also provide wifi connectivity, with the possibility of exchanging wifi for people's attention by offering incentives to those passing by to shop in nearby outlets, and can read what people are doing online while connected to their signal. In other cities, smart lampposts have been deployed that can pick up mobile phone traffic in the area.

Like CCTV, these lampposts offer the potential to light areas more efficiently and responsively. Along with this they also offer various possibilities for surveillance and monitoring in public space, and also for commercialising and monetising people's presence in that space. A range of types of sensor can be added to smart lampposts, allowing them to sense foot traffic, behaviour, objects and activity. This 'Christmas tree' characteristic allows city authorities to develop and change their function over time in response to perceptions of risk, possibility and efficiency.

3.2. Privacy, responsibility and accountability

3.2.1 Distributed data

One of the reasons that big data presents new problems in terms of governance is because it is distributed: it exists across databases, applications, owners and permissions. In contrast to this, the notion of privacy self-management[23] as put forward in various data protection instruments holds that we must be able to take responsibility for the data we emit, that we must be able to own it, to check it for mistakes, and to revoke the privilege of using it if our relationship with providers is abused. We can contrast this, however, with the real world in which 'our' data is processed by many layers of actors in parallel, consecutively and often simultaneously: government systems, communications providers, app developers, data intermediaries, banks and e-commerce firms, employers, and our own networks, to name but a few. In the real world, our data changes hands many times a second and huge international data intermediaries hold thousands of data points on each one of us, including private information about health and behaviour. As senior Microsoft official Craig Mundie[24] has said, 'today, there is simply so much data being collected, in so many ways, that it is practically impossible to give people a meaningful way to keep track of all the information about them that exists out there, much less to consent to its collection in the first place.'

When we asked our interviewees and focus group participants about how they conceptualised and handled their digital privacy, most realised that the data they emitted went in many different directions and was collected and used by a variety of public and private bodies. This gave rise to an awareness that data had

[23] Solove, D. J. (2013) Privacy Self-Management and the Consent Dilemma'. *Harvard Law Review*, 126, 1880.
[24] Mundie, C. (2014). Privacy pragmatism. Foreign Affairs, March/April. http://www.foreignaffairs.com/articles/140741/craigmundie/privacypragmatism

become untrackable and hard to audit, and that no one had either the time or capacity to engage in privacy self-management. As the discussion progressed, the participants mostly arrived at the realisation that they did not have the necessary information either, making them highly reliant on authorities for preserving their digital privacy.

The city has to manage privacy for the projects it conducts that use data about people's movement and activities. However, this can be complicated because when the devices and the data are distributed, it is hard to keep the responsibility and accountability from becoming distributed. One important reason that data is distributed is that city administration and operation are also distributed functions. The political administration of the city (the mayor's office and city council) is separate from the city's law-enforcement apparatus, from the whole bureaucracy providing public services, and from infrastructure departments such as traffic and building permissions. Unless a city has internal, centralised capacity to do the data science and statistics necessary to use all the data that becomes available (as for example New York's mayor's office has established),[25] it will have to establish partnerships with commercial firms to access and analyse data, meaning that there is seldom a single actor in charge of a particular research or application using digital data.

3.2.2. Distributed databases

These different uses of data by different parts of the city government also lead to distributed datasets, and creates the potential for any problems that come up with storing, reusing or deleting data to be fragmented amongst departments. One example of this is traffic data: if the traffic management service (Dienst Infrastructuur Verkeer) wants to know how a new piece of infrastructure such as a bridge or a tunnel is affecting traffic flows, traffic managers can study those flows using a system of highway Bluetooth sensors, independent from the traffic information flowing through national systems such as automated number-plate recognition.

Distributed systems for collecting and processing digital data also raise the question of data ownership. The city often does not own data that is used for urban governance – and as more complex data sources and analytics are increasingly used, the extent to which the city controls the data is in fact diminishing because it does not have the internal capacity to do the analytics. In the case of the Rode Loper project, for instance, the mobile phone data used was handled under a licence where the firm that collected it gave access to the municipality for a specific purpose, but kept the data in its own database.

Amongst our respondents it was common to have experienced their data flowing across categories. All knew that law enforcement could get access (under specific conditions) to almost any data about them, but were unclear as to how other interdepartmental and inter-institutional permissions for sharing and using data worked. Some had experienced serious problems: one interviewee recounted being denied a mortgage due to a particular health condition that they had only disclosed to health authorities. Some people we spoke to, however, were used to their data flowing across departments, notably where sex workers (who have to register with city authorities in order to work) find their registration can also be accessed by highway police, who identify their vehicles and stop them for search and questioning (sex workers focus group, 10.12.2015). Registration as a sex worker also leads to housing problems: the Chamber of Commerce keeps a public database of all those who register a business, with names and addresses of those businesses available on the web. For those whose business is sex work, this involves publishing their personal details. This can make it hard to find housing, as landlords are unwilling to rent to sex workers. A new proposed law would also make it compulsory for those who work from home to put up a sign on their door to indicate their profession, effectively classifying their home as a brothel, and would demand infrastructural changes that would effectively identify a sex worker's home as public space:

[25] The NYC Mayor's Office of Data Analytics (MODA): http://www1.nyc.gov/site/analytics/index.page

> (participant 1): '[according to the new law] you need to keep to the same standards as a brothel owner, for instance you need to have a fire escape and things like that...
>
> (participant 2): Things you wouldn't need if you'd run a beauty salon from home for example.'
>
> (sex workers focus group, 10.12.2015)

Being in a government database as a sex worker, then, leads to a mixing-up of private and public space. The measures are designed to fight trafficking by creating visibility and clear guidelines to distinguish those working freely from those being forced, but this aim also involves using public registration details to check on private space. Participants in this focus group reported that it was common for their homes to be searched by police as if they were brothels:

> 'It's the same thing when the police come in to do their regular check-ups of your work space, whether you work from home or in a brothel. For example, they'll ask you about your personal relationships, they'll ask you about your ID, they'll ask about your children. We have police entering the houses of sex-workers, entering those houses claiming that they have the right to look through the closet, looking at the clothes, saying these are clothes of men, do you have a partner? Where is he? Where is your kid when you work? What kind of sex do you have? Is your personal sex-life satisfying?...'
>
> (sex workers focus group, 10.12.2015)

Although sex workers use public and private space in ways that inevitably have implications for regulation, this characteristic is common to many different groups. Just a few examples include commercial traders, taxi drivers, police, performers, builders, tour guides and many others. The case of sex work is interesting for this study because it provides insights into what can be done with data when a group is judged risky in some way, and especially because it shows how digital data, as a tool for knowing and regulating a population, tends to act as an equaliser, erasing differences and nuance in terms of behaviour, regulatory status, and activities. All these groups may be individuals or incorporated, may both perform valuable services as well as potentially intersect with legal issues or risks to the public, and may at the same time have both private and public personae. From the case of sex workers we can learn how many groups may find themselves positioned with regard to the new data technologies in city life. For example, the negative effects of public registration impinging on private space leads sex workers to create pseudonyms, and even pseudonymous spaces, in order to force a distinction between their public and private identities – much as many individuals choose to have multiple online identifies to serve different functions in their lives. Yet pseudonymity is a problem for registration, which is designed to pin down the individual and their function in the city in ways that can be inconvenient or even invasive for those who do not want private aspects of their identity becoming public:

> 'I asked the city of Amsterdam to keep my data secret or not make it public, and I hired a flexplek (a one day a week work spot) in an office, so when you look at my company at the KvK [chamber of commerce] you only see my full name (for fuck's sake) and my office address, which is better. It's hugely problematic. I try not to [become identifiable], I know that with the tiniest bit of techsavyness its still useless but I try to go for some anonymity. I am not online as my wallet-name anymore. My only online presence is my work presence and my activist presence.'
>
> (sex workers focus group, 10.12.2015)

The sex workers' experiences demonstrate how profiling and monitoring performed for purposes of care (combating forced sex work) can easily be experienced as control. Being flagged in a government database as risky – or, equally, at risk – leads to more prolific data-sharing across governmental departments and a greater

likelihood of intervention by authorities in one's space.[26] From public health, public safety and anti-trafficking controls come registration and official control over working and home space, and these in turn bring risks for the individuals being tracked and registered. There is also the question of how private data analytics (hacking) may intersect with public data collection and tracking, and how a too-narrow conception of 'the group' or of 'risk' may backfire in terms of rights.[27]

The idea of risk plays out in many different ways in the smart city: risks of inefficiency, risks to public safety, risks to public health. The idea of crowds and busy crowded places as inherently risky is also central to many 'living lab' experiments and research projects such as the Rode Loper crowd observation project and the smart lamppost project. Technologies that capture people's electronic signals and monitor their presence can also become part of systems that predict risk – who is behaving in an unusual way, who is loitering, who is gathering together in a way that might indicate unrest? Such technologies have the potential to radically change public space, and it is currently unclear how to go about democratising them and making it possible to opt out of digital visibility. One expert interviewed mentioned that

> 'It's interesting, there seems to be a move from the living lab from the sort of experimental context to do things permanently, where you get some kind of consent structure on the living lab experience and nothing for the larger projects, is that what you're seeing as well?] (researcher, interviewed 12.3.2015)

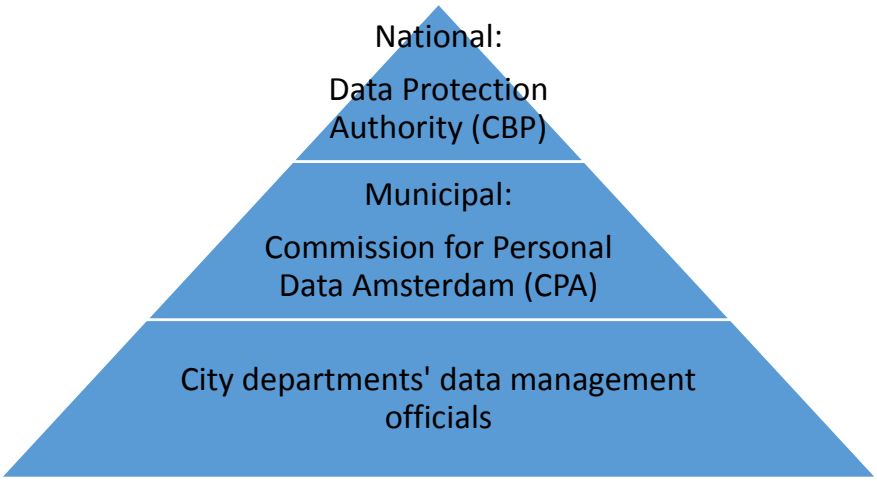
3.2.3. Invisible data infrastructures

Data infrastructures are governed both by those who set them up and use them, and by whoever is in charge of regulating them in terms of security and privacy. The collection of digital data in Amsterdam is managed (or potentially regulated) in several different ways (see Figure 3): those wishing to conduct research or commercial projects that involve collecting digital data on residents must first check with the relevant department, and gain permission from officials there to collect and process the data. There is also an independent commission on privacy and personal data, the Commissie Persoonsgegevens Amsterdam (the Amsterdam Commission for Personal Data), whose job it is to review any risks to data protection in the municipal context and to inform the city council and administration. The CPA is supposed to discuss issues and complaints submitted by, among others, the city ombudsman or the Integrity Office (Bureau Integriteit) of the city, but also by individual Amsterdam residents. There is also a national data protection authority, the CBP, but this office generally deals with national-level issues of data misuse. It does not deal with anonymised or de-identified data of the kind frequently used in smart city research-and-development projects because these are exempt from data protection law – although, importantly, location data is becoming a grey area for EU data protection regulation now that it is recognised that one's location can be both identifying and can convey details about a person's activities and personal characteristics.

[26]See Keymolen, E., & Broeders, D. (2011). Innocence lost: Care and control in Dutch digital youth care. *British Journal of Social Work*, available at: https://www.researchgate.net/profile/Dennis_Broeders/publication/277524746_Innocence_Lost_Care_and_Control_in_Dutch_Digital_Youth_Care/links/55b74f7508ae092e96570f0e.pdf

[27] See, for example, http://www.theestablishment.co/2016/01/28/why-using-hacking-to-eradicate-sex-trafficking-is-dangerous/

Figure 3. Data protection at city and national level



The city's data authority (the CPA) is a public committee designed as a check-and-balance structure for the data collection and processing permissions issued by city departments. It meets regularly, the records of its discussions are available on the city's website,[28] and it is a public-facing institution. However, our research found that today's digital data collection and use practices present a challenge to the CPA. Like the national authority, it cannot be expected to proactively find every instance of data use that may be problematic, but is reliant on complaints from the public or from officials. This makes it difficult for the Commission to have a preemptive function in guarding against potential harm from projects still being conceptualised. Nor can it operate on the scale on which data is now being collected and used, since it meets once a month for three hours, and its members are retired volunteers with expertise in the field of digital data and privacy. There is also a challenge in terms of people's awareness of the CPA: not one person we spoke to for this research had heard of the Commissie Persoonsgegevens Amsterdam, although one researcher had a project that, as it later turned out, had been approved by the CPA. In this case the permission had been handled by one of the junior researchers attached to the project so that the project leaders had not been part of the permission process.

Beyond the challenges of scope and public awareness to do with the CPA, there are more structural (and spatial) obstacles to its function. In the case of monitoring and data collection that take place in privately owned space, such as for example at a football match or an arena concert, only the organisers' permission is necessary and therefore neither departmental authorities nor the CPA would be consulted. Urban mobility researchers interviewed noted that

> '…it depends. When you monitor at an event and the event is held at a confined spot, by an event organiser, you don't have to go through them [the city authorities] because you are talking directly to the event organisers who do all the permissions.' (academic transport researcher, interviewed 18.3.2015)

There are therefore many projects where no permission needs to be sought from public authorities, nor would participants in an event know they were being monitored.

Events where the city government is a partner in the monitoring, however, are starting to include notification to people that technology is being used that may capture their images or movements. For example SAIL, a maritime event held once every five years in Amsterdam where tall ships come from all over the world to dock in the city, is now subject to the same kind of research for crowd management as took place in the Rode Loper

[28] https://www.amsterdam.nl/gemeente/organisatie/overige/adviesraden/commissie/

project. In 2015 the authorities put up billboards to let people know that remote monitoring of crowds and movements was occurring (Figure 4).

Figure 4. Billboard posted in the Amsterdam metro during the 2015 SAIL event

Other public offices are seeking ways to apply ethical principles to new data collection and analysis methods. A researcher at the Environment Bureau (Planbureau voor de Leefomgeving, or PBL) suggested that a way needed to be found to apply existing survey ethics, where people give their consent in person and for a specific use of the information collected, to big data:

> Interviewer: 'So you take the survey-based model for permissions and apply it to big data?
>
> Researcher: Yes, probably, that would be my first instinct, but we are a small institution so we wouldn't want to have to deal with sixteen million Dutch people, or seventeen, however many.' (interview with PBL researchers, 17.3.2015)

In this case, existing survey ethics are the main method available for seeking informed consent from people when collecting and analysing data about them, and the method envisaged by European data protection law if those data include personal information.[29] However, as the interviewee notes, on a national (or even city) scale this becomes next to impossible to manage. Furthermore, as data is increasingly linked and merged in order to discover new insights, it is increasingly difficult to predict what the data people generate will be used for, or how those uses and combinations of data might change over time. This poses an even bigger problem for consent, since – at least in the easy cases where data is collected directly rather than remotely – people may be informed about one use of their data but not the whole range of future possibilities.

[29] EU Directive 95/46/ec, which is scheduled to be replaced in 2017 by the General Data Protection Regulation (GDPR)

In one focus group, with immigrants from within and outside the EU, the issue of the right to be forgotten came up. Participants were not aware of the challenge of exiting from the various databases in which their details were stored, even in cases with official data where they knew they were in the database.

> 'On the governmental side, things like the criminal record. On the one hand I understand that some people say you have to keep the criminal record and there is no way to delete it because the people have to know what you did wrong but there are so many cases where I can imagine that the criminal record stops you from getting a job or stuff like that because you have to show it... often it's required to show a criminal record. If somebody had some troubles with marijuana in his youth. He will have this for his whole life, it's kind of a CV then and that's not fair.'
>
> (Immigrants focus group, 2.12.15)

Criminal records are an example of a data use that would seem to be fairly black-and-white, but this issue came up in a larger discussion of how to exit databases involving social media, internet search records and official records of all kinds. Although participants were clearly aware that citizens do not have the right to withdraw from the records, they identified the problem of control in the datasphere as a whole, seeing it as a complex mix of rights, duties and simple capture of personal data.

This complex interplay of citizenship, rights and duties is heightened by processes of digitisation. There was a high-profile 'right to be forgotten' case against Google in 2014, where it was established that European citizens had the right to have internet search results relating to outdated or unfair material deleted. As was argued in that case, an internet search firm has a different kind of function than a newspaper because the information it provides is more easily found and disseminated than traditional hard-copy newspaper records, or other analogue archives. Similarly, when one's name is entered in a city database it is differently searchable and potentially differently visible than it would have been before widespread digitisation.

The experience of one Amsterdam resident in the focus group of non-smartphone users demonstrates how not being able to exit the database can impact on one's experience of citizenship and rights: she was the victim of an assault, which she reported to the police. However, by doing so, she found that her name and address would become part of the record of the alleged crime and that the accused person would also have access to them in order to build his defence. This meant that by reporting a crime, the victim had to put herself at risk of reprisal. Similarly to the problem of the sex worker who has to register in a public database in order to work legally, the imperative to be responsible and accountable is here explicitly at odds with the person's privacy and personal security. In the era of big data, public authorities will increasingly be confronted by the tension between people's identities as private citizens, with the right to autonomy and privacy, and their identities as public citizens, whose digital selves are continually captured and replicated in databases and used to operate and shape the city. Rights to our data are not simple to apportion because the more detailed the data that is held about us, the smaller the gap between our digital and real selves.

3.3. Participation in innovation

The evolution of the various smart city projects in Amsterdam (and in cities worldwide) is marked by an emphasis on innovation but also on collaboration. Commercial firms, startups, individual developers and increasingly the citizen public are invited and enlisted to participate in inventing and creating the smart city. Responsiveness, feedback and public involvement are frequent hallmarks of planned projects. We were interested in how this discourse of citizen involvement and participation interacted with the increasing use of digital data from multiple sources. The actual data collection and analytics elements of smart city projects were fairly opaque to most of our non-expert interviewees: they felt part of a general data market, including the

data they volunteered to municipal departments, but they did not feel included in the planning or execution of smart city projects and research in general.

Box 4: Social Glass

The Social Glass project (www.social-glass.org), developed by a team of scientists at TU Delft, aims to provide a real-time picture of the dynamics of city life. The project monitors streams of social media such as Twitter and Instagram, linking it with statistical data and the Open Linked Data Cloud. The developers aim to provide policy makers and city authorities with better knowledge about city operations, so that they can respond to events in real time. The idea is that if public opinion and experience can be merged with larger-scale operational and socio-economic data the city will become more manageable and more governable. The tool is also based on the assumption that the sentiment expressed in social media posts will provide qualitative information that will balance out the automated collection of sensor data from cameras and GPS trackers.

The project was tested during the SAIL2015 event in collaboration with the Institute for Advanced Metropolitan Solutions (AMS Institute). The developers observed 60,000 social media users during the event.

The aim of the team is to develop a tool that can go beyond real-time analytics to anticipate and respond to long-term trends, and that can also be used by citizens to better understand their city.

3.3.1 Bias in citizen involvement

When selecting our focus groups we looked for a mixture of average citizens and those who are likely to be marginalised by digitisation. Across these groups, we found that the respondents involved in this study were unaware of who was organising smart city projects in Amsterdam, how they might have input, or even how to find out about what was happening. Even the highly-connected people, including technology developers, that we spoke to were sceptical about the link between digital urban smartness and participation by ordinary citizens. The city has made an effort to involve those who are interested: the 'Amsterdam' smart city website is a resource for many of the public-facing projects taking place under the banner of smartness and digitisation,[30] and amongst them there are many that seek to involve citizens – a hackathon, neighbourhood regeneration projects involving digital networks, an environmental metrics project, a bootcamp for digital entrepreneurs, sustainable energy projects, a visitor feedback system for museums, and a digital network for sharing cars in the city. The projects are wide-ranging, both geographically and in terms of content and aims, and promote a vision of a public-facing, engaging process of developing the smart city. Yet, apart from a couple of respondents who were professionally connected to smart city research or projects, the people we spoke to were unaware of them.

There is also an inevitable technological bias in the projects that make up the smart city's current stage of evolution: they tend to be directed toward younger, more educated, more technologically aware city residents. The images available show little diversity in terms of class, ethnicity or age. Although many of the projects aim to involve whole neighbourhoods, it proves to be difficult to present that aim clearly through the lens of technology and innovation. It has been suggested before that inclusivity suffers where cities aim for

[30] http://amsterdamsmartcity.com/projects?lang=nl

smartness,[31] and that by its nature, the digitised city has trouble engaging with those who are uneasy with, or excluded from, the digital world. Another natural bias is that the city's digital world tends to be skewed towards commercial participation – one of our expert interviewees who works in transport analytics noted that the city government was keeping citizen data under city auspices in every way possible, to the point of rejecting price-saving collaborations where they would mean that commercial parties made money from the data collected as part of crowd management research. Despite this effort, the same technologies used to research crowd dynamics are used by retail firms to track customers, so that data collection efforts in the public and private sphere increasingly parallel each other, creating similar but diverging data flows, one commercial and one public-sector.

The public face of Amsterdam's smart city projects is designed to promote a sense of possibility for participation, and this certainly exists. However, respondents generally voiced a desire to be able to resist and exit what they saw as 'the system' of data collection and use rather than to participate in it. There were several, particularly amongst the student respondents, who were enthusiastic about the possibilities of the smart city and digital data collection and use, but they were in a small minority compared to those who wished to find ways to resist datafication. Many saw the smart city as a neoliberal, modernist project – probably because of the central role of private firms – and felt that their role as citizens was getting lost amidst a rising tide of digitisation.

3.3.2. Citizen science: measuring one's environment

Amongst the category of initiatives that could be described as citizen science – collecting data directly on the environment and on neighbourhood issues – an implicit assumption can be found that empowerment is about getting people involved in participatory open data collection and creating more spatial digital data on citizen-relevant issues. Often this manifests as organising a day where people walk into the street with their smartphone and collect data about how many instances there are of a particular category of interest. For example, counting how many people are walking and what they are doing whilst walking, and whether that affects how people talk on the street, as a proxy measure for neighbourhood friendliness. A notable example is the Measuring Amsterdam event from the Hogeschool van Amsterdam's Citizen's Data Lab,[32] which was a moment to focus on citizen empowerment through participatory open data collection. Whilst discussion did touch on issues of inclusion in bottom-up initiatives, the essence of the day was crowdsourced images of complaints on the street and a digitally mediated voting system for proposed solutions – both of which leave out those who are not digitally aware and active.

Notwithstanding the possible methodological problems in using only a particular segment of the population to report, and only a particular slice in time to collect data about the way the street is, one expert interviewee noted that whilst more data does give broader tools to discuss neighbourhood issues, participatory open data collection is not empowering in the way we often imagine the emancipatory potential of the internet.

> 'It's a very normative take on citizen empowerment, [...] that the society is an energetic society, and if we provide the society with the proper infrastructures people can empower themselves. It can be institutional political infrastructures, but also information from data. Once you know how data flows or how materials and people are flowing, you can maybe change the idea on how the world around you functions and how you would like it to function. A lot of it is about making things transparent for the larger audience, which I think is kind of a more general idea behind the development of infographics. [...] Sometimes it's a fairly technocratic narrative that goes some way along the lines of

[31] See, for example, Adam Greenfield's *Against the Smart City* (http://www.architectural-review.com/archive/reviews/that-smarts-against-the-smart-city/8667075.fullarticle)
[32] http://www.measuringamsterdam.nl/

> you know, we have this great data visualisation of how the impact of particular water related policy in the municipal area and if we present this, that will enable democratic decision making.'
>
> (Interview with researcher at PBL, 17.3.2015)

The smart city's development, then, needs to find a way to invite in other perspectives beyond the technocratic. Otherwise it is in danger of a chicken-and-egg situation where the people included tend to already be the highly educated who are excited by smartphone apps, deepening the wedge between in- and excluded. There is the possibility of a virtuous, rather than vicious circle, however, where people participate voluntarily in the datasphere in order to increase the accuracy of big datasets:

> 'You know, I'm much less scared [of data maximisation] than I thought I was. I'm scared of being quarantined in a data-poor environment. I don't know if I'm scared of city planning in a data poor environment or in a data rich environment. I don't know which is weirder. Probably the poorer environment is again more problematic, because yeah, it's less accurate. … so this might also be the incentive for people that they know if they give away their data, that they're at least privileged in some respects, know that companies and municipalities etcetera have the correct data and they're not on a false positive list of some kind.'
>
> (Researcher, interviewed 12.3.2015)

Taking this circular dynamic further, interviewees who were themselves technology developers also identified a feedback loop between personal visibility and functionality in the smart environment, explaining that the online economy, particularly around open source software, demands personal openness and visibility in return for credibility and uptake of innovation:

> 'I come from an open-source background and in an open source background it's very common to put everything online… you create and people get to know you know from what you are doing... so not telling about what you're doing but doing it. But as a company you need to promote yourself a little bit more than just put online the code you're making.., because people that take the decision whether to hire you or not to hire you or not don't always understand the code. So you need to tell more about yourself than just putting your code online.'
>
> (Freelance developer, interviewed 26.6.2015)

In general the more technically involved people we interviewed were more willing to trade data on their activities, location and movement for innovations that would increase the ease with which they could move through the city, but also, by extension, participate in its life. There was an uneasy bargain experienced, however, between using an innovation and participating in developing it by donating data – often beyond the point that people felt comfortable.

> [focus group participant 1]: 'I think we accept to give up our data when it really aids us in daily living, like stuff like Google Now for instance. It says like 'leave now for this and that meeting', I find that aggregating these data sources [your agenda, if there are traffic jams on your route, weather, etc.] I find that this is really smart, it's really helpful to have that stuff and I am willing to give on privacy for these use-cases.
>
> [interviewer]: Do you make decisions about who you trust to have your data, or is it just that it is out there and there is nothing you can do?

> [participant 1]: If you are given the choice [about who to trust with your data] then you do think about it. At least I do think about it. [participant 2 agrees]
>
> [participant 3]: Yes. But for a lot of services there's no alternative. It's so well integrated with, for instance your phone, that you can't do without them you have to use them. You are forced.'
>
> (technology developers focus group, 3.12.2015)

One main finding that came out of this range of discussions about innovation and measuring one's environment is the paradox that when you measure the environment or your neighbourhood, you also measure yourself. In order to participate in processes of digital innovation and smartness, people must embrace their visibility both for the sake of the technology's development, and for the accuracy of the data it will produce. The smart city is, by its nature, a collective endeavour, even though it is biased towards top-down processes that privilege the technologically adept.

The other issue that came out most strongly was people's nervousness and anxiety about being surveilled in public space, and of their everyday activities becoming part of the global data market. From the discussions it appeared that this feeling stemmed less from concrete evidence of surveillance than from the knowledge that all their online activities were monitored and monetised, that data breaches by trusted authorities and firms are frequent, and that they signed away their privacy whenever they accepted a user agreement. These elements combined to create a sense of distrust towards data-gathering in general, except for the most basic functions of citizenship which they accepted had remained much the same, but had been digitised. The anxiety and fear generated by ubiquitous data collection seem to strongly influence people's feelings about smart city projects and infrastructures, so that they feel less a sense of possibility and more resignation to being data subjects. Finding ways to gain public confidence appears to be a central challenge for the smart city movement.

4. Ways forward: integrating the citizen into the geo-information infrastructure

4.1. How can the city regulate the urban datasphere?

Even though it only controls a small portion of the data that circulates in the massive global datasphere, the city has great power to establish and enforce good practices to do with data. City authorities have the power to procure new technologies and infrastructure, and thus to influence the development of technology and databases directly. They also have power over education (identified by many focus group participants as a critical missing link in promoting greater public data awareness and engagement), because they are funders, and can therefore influence – though not dictate – data responsibility.[33] Cities also have the ability to stimulate the regional economy, and thus can influence a broad range of types of regional development and innovation.

The scenario-building element of our research in combination with the focus groups' responses to those scenarios suggest that cities can also manage their partnerships with the private sector in ways that promote a more inclusive and democratic datasphere, in ways not limited to the procurement of technological systems. They can create contracts that ensure city authorities have full access to data stemming from the provision of public services; they can create public consultations around the development of systems and infrastructures,

[33] For example, regarding connecting children to the data market though school laptop initiatives: https://www.eff.org/studentprivacy-casestudy?from=student-privacy

bringing the public into the decisionmaking process about what the city needs, but also how it should function and what kind of information they are prepared to contribute. They can also promote public discussions about consent, privacy and autonomy with regard to digital data – our findings suggest that the direct personal connections formed through city services such as welfare and employment, permits and citizen registration offer an opportunity to break down barriers of misunderstanding and lack of information, and to both inform and invite feedback about data use and sharing. These points of contact, however, are being eroded by digitisation. For example, at the time of writing changes are planned that will digitise interactions between the city and welfare applicants.[34]

One concrete opportunity presented by the city's position as an alternative to the commercial datasphere is not to build a future as a data broker between citizens and private sector contractors, but instead to innovate by preserving its advantage as a closed loop in terms of data processing and sharing. The city has so far resisted allowing commercial firms to sell on data gathered as a result of smart city research and development projects. However, apps are becoming the shortest route to connecting people to services, and apps are built on a model where developers amass all the data from their use. However, the city has a monopsony[35] on the particular services it provides – for example public infrastructure, welfare and registration processes – and could potentially develop apps of its own that would not share data with outside parties. The app economy could thus develop into a win-win for the public in terms of urban service provision, providing personalised service and allowing real-time feedback while also preserving privacy. Another opportunity for the city is to consider retaining – or even increasing - points of face-to-face encounter between government and citizens as digitisation progresses. As focus group participants have pointed out, it is difficult, if not impossible, to personally speak to municipal civil servants in relation to most operations of urban citizenship.

One major issue to be solved if the datafied, smart city is to serve all its people rather than just the most technologically integrated is that of tailoring data collection and sharing to context. Philosopher Helen Nissenbaum has written[36] about how the way to create and preserve trust amongst technology users is to keep data in contextual streams – for instance, if I give a provider my personal details in order to ensure I receive my pension or appropriate healthcare, I should not then receive marketing based on that information from commercial companies. Official data collection for purposes of registration, however, tends to mass people together in ways that violate the contextual integrity of their information. A registration process that is unproblematic for a hairdresser may be experienced as abusive by a sex worker, since what is positive publicity for one is a violation for the other. In fact, any registration system that applies the same rules to individual operators as to collective firms is likely to experience problems in terms of accurate information, as people will employ measures to become less visible and decrease the potential for problems, and those measures will then decrease the reliability of the database. A city where everyone is forced to be equally visible would be a profoundly dysfunctional place to live. As one of our interviewees suggested, anyone who thinks they have nothing to hide and therefore doesn't need control over their visibility should try removing their bedroom curtains. Instead, one suggestion that arose in our focus groups was that rather than creating one solution to fit all, the government could develop separate online spaces or classifications to cater to particular data privacy needs.

[34] Personal conversation between project researcher and Amsterdam welfare digitisation project manager, 17.6.2016

[35] A condition where a single buyer is so powerful they can determine what is produced.

[36] http://www.theatlantic.com/technology/archive/2012/03/the-philosopher-whose-fingerprints-are-all-over-the-ftcs-new-approach-to-privacy/254365/

4.2. Citizen insecurities and alternative visions

Historian Jan Holvast, interviewed for this project, explained how the Dutch census boycott of 1970 occurred.[37] The census that year was to be computerised, with each person allocated a number and their details collated using punchcards. The questionnaire delved into private issues such as personal views, income and handicaps, and participation was made compulsory, with noncompliance punishable with either a significant fine or a prison sentence. All these factors reminded people of the Nazi occupation, when IBM's automated tabulation systems were used on census data to identify, victimise and deport Jewish and other undesired citizens. People felt threatened, but also protective of marginalised groups and in particular undocumented migrants. In response, there rose up a popular resistance to the census that resulted in its abolition and replacement by a municipal register, which persists to this day.

Holvast's history of the boycott is relevant to this report for several reasons. First, it demonstrates the deeply embedded insecurity about being counted, categorised and recorded that is being reawakened by datafication. This insecurity was voiced by all our interviewees without exception, including technology developers, data scientists, academic researchers involved in smart city projects and city officials themselves. Second, the main group that drove the census protest was in Amsterdam, where people both rejected the echoes of the Nazi Occupation, but also felt solidarity with those who would be disadvantaged or marginalised by the data collection methods proposed. Finally, the story is important because it echoes our alternative smart city scenario, 'Anonydam', a less individually visible but more interdependent urban population of the kind envisaged by Adam Greenfield in his UrbanScale project.[38]

It is possible to question how socially interdependent people are willing to be today, in comparison to the Amsterdam community of 1970. Holvast notes that 'the most important difference between now and then is that then we were talking about *our* privacy, and now we are talking about *my* privacy'. This was also a finding of our research: people were concerned about privacy mainly in the context of their individual relationship with their online world, or with the city, rather than about the privacy of their families or their networks. This suggests that what constitutes good data governance has evolved since 1970. Instead of protecting the most vulnerable, people are interested in a system that protects the citizen user, the contributor of data – yet however individualistic that idea, it can be argued that a system that protects one individual can be made to protect all individuals. A more responsive system benefits everyone, and is also potentially more inclusive. But what would such an inclusive urban data governance structure look like? It is hard to imagine a system that is innovative and takes advantage of current technological possibilities, but that is also accessible and friendly to the elderly, the marginalised, the less educated and the lower-income, and that allows a certain degree of opting out – at least in terms of mass data collection. Such an inclusive governance structure would also have to deal with the less visible forms of data collection and use, of which people are mostly unaware.

There are two different conceptualisations of informational privacy that emerge both from the literature and from this study: one is based on a model of private property where the individual keeps control of information about his or herself. This vision of privacy involves keeping the public sphere out of the private sphere. The second vision is about preserving the integrity of one's personal infosphere, and thus keeping a personal space within which relationships, sharing and autonomy can be created and preserved. The second relates most closely to the findings of this research. Both models of privacy have a spatial aspect: the personal sphere is conceived as a space within which one functions, and within which one can choose to accept or reject contact of various kinds. We do not argue that spatial data deserves special consideration in terms of urban data governance, but that there is a spatial aspect both to what our respondents wish to see created – data infrastructures that serve people as citizens and that open up the city to its people, rather than just opening up

[37] http://www.npogeschiedenis.nl/andere-tijden/afleveringen/2011-2012/Volkstelling.html
[38] http://urbanscale.org/about/

the people to the city – and similarly a spatial aspect to what they wish to preserve, whether that space is visualised as the home or the self. Both our movements and the places where we stay still tell almost as much about us as our internet search history, but are significantly less well protected.

4.3. What capacity needs to be built by cities?

We stated at the start of this report that we are interested in future technological developments, and in the emerging problems of urban data governance, informed by what can be seen in practice today. Thus we return to our central questions: first, what organisational and governance structures are necessary for a sound and innovative spatial data infrastructure in Amsterdam? And second, how should the city address the needs, privacy rights and responsibilities of citizens who create and use spatial data? In response to the first, our findings suggest that the city's management of its own data is an important capacity to build. Cities that have followed this model find themselves with more options with regard to their engagement with the private sector, and also potentially with better connections to their citizens and more opportunities to build trust around data collection and use. In response to the second question, there is an emerging democratic deficit with regard to the way data is collected and used. Data is essential to democratic representation: the social contract entails people making themselves governable in return for good governance, and visibility is a necessary component of governance. However, the more efficiently systems create visibility, the more democratically controlled they must be. One way to operationalise this is for the city to position itself as a trusted intermediary, developing its own data scientific agenda and fostering data scientific innovation within its own bureaucracy. In the internal environment, the incentives for monetising data are minimised and those for creating efficient, data-conservative services are maximised. Creating this new space for data science may be the most effective way to involve people and build trust in those digital operations that are best performed remotely, such as data analytics for parking allocation, energy management or fire safety.

What is necessary to achieve these ideals? First, technical capacity within the city administration, as well as partnerships with innovators outside it need to be built. Our professional developer interviewees had relatively little faith in governmental technical capacity, and therefore felt resigned to the city's data eventually forming part of the global data market. However, this is not inevitable – there is also the option to prioritise building internal capacity to balance out the participation of contractors. If city managers can better audit what data is produced and how it is processed, they will be better positioned both to use it and to act as a responsible intermediary for the citizens who produce it.

This leads back to the human dimension of data governance: building a more inclusive datasphere also means engaging with some potentially uncomfortable political debates about issues such as profiling, what should trigger data-sharing across contexts, and – perhaps most importantly – how to include currently marginalised groups in the discussion about data governance. One example of this problem is illustrated by the resistance we encountered from city gatekeepers when we tried to engage with one particular group – sex workers – about their data. At one point we were told we could not invite a group to discuss digital data because they would not understand our research. 'They will not understand these questions,' a city official told us, 'they are not sex workers for nothing.' In fact, the sex workers with whom we were able to engage on these questions were some of the most educated and articulate people we met in the course of the project, and provided some of the best-articulated insights of any group.

Another part of the human dimension is strengthening intermediary institutions, such as the Commissie Persoonsgegevens Amsterdam but also citizen organisations such as the Waag Society, Bits of Freedom and activists for participatory data such as Stichting GR1P.[39] These institutions perform the valuable function of

[39] http://gr1p.org/

centralising information about what works and what is problematic, and act as a counterbalance to the distributed data infrastructures that characterise the urban governance challenge. The CPA, like the national data protection authority the CBP, is overstretched in terms of its mandate. The Commission is run and staffed by people with significant expertise who are committed to their role in safeguarding privacy in the city. However, the challenges of today's smart city programs seem to require a huge increase in the amount of attention appropriate to assessing the privacy concerns raised by datafication.

Finally, the rules may also need to change if information infrastructures are to answer the needs of the smart city. For example, the rules for data permissions should be able to cross between public and privately-managed space – from the street to the football stadium, or the metro station to the nightclub or even brothel – just as individuals and their data do. Having different sets of rules for the collection of data in publicly versus privately managed space is likely to lead both to the inappropriate collection and use of data, and to protest if data is known to be mishandled. This will become particularly problematic to manage as internet connectivity increasingly mingles public and privately owned space – the next generation of wifi hotspots will be small devices attached to city infrastructure such as lampposts and billboards, but also houses, office buildings and shops, with the aim of providing a network of connections that people can move between seamlessly, without signing on to each separate source. Data, like individuals, will then increasingly cross public/private boundaries in real time. This means that individuals will either carry a set of permissions with them embedded in the data they emit, or that overarching regulation will determine what data can be collected and how it can be used. The likelihood, in fact, is that these two approaches will mingle and that if cities want to guard and preserve people's right not to be tracked and read as they pass through urban space, they will have to become actively involved in regulating and intervening to create the kind of urban datasphere that serves everyone's interests rather than channelling power and data to large-scale commercial interests. Currently, the default model is distributed decisionmaking, but more centralised awareness and command over data may have to be developed to keep the datasphere fair and democratic.

Related to this, cities will benefit in future from thinking beyond the compliance paradigm (what is legal) in terms of data collection and sharing. For private contractors involved in city datafication projects, privacy has become an issue of compliance (assuming that the use of de-identified data does not constitute tracking, for instance) and conformity. In contrast, a successful data governance infrastructure in the future will be less focused on compliance and more on equity and understanding. A citizen-centred framework for data management, rather than a solely commercial one based on legal compliance, is more likely to address the moral and emotional aspects of data space identified in this report. It will also open more potential space for democratic accountability and debate. Our research shows that rather than being uninformed or apathetic about their data, people are engaged but frustrated. They strategise, they minimise, but (as Joseph Turow points out in his recent study of consumer perceptions of data science[40]) they become resigned when they are on the losing end of data misuse, and that resignation looks like (and is taken as) apathy and contradictory behaviour by those whose role is to monetise their data.

The development of an integrated geo-information infrastructure requires a rethinking of the relationship between privacy and space. This is not only important conceptually, but also for the design of the above-mentioned government-citizen dialogue and the development of regulatory frameworks. On one hand, we found that it is especially location data emitted more or less involuntarily via people's use of GPS-enabled devices, or emitted more consciously via social media, that raises the feeling of infringement of privacy. On the other hand we also see a new regime emerging, especially important in the case of geo-information infrastructure, where the categorisation of space into surveilled and unsurveilled no longer makes sense from an individual perspective, because it is possible to digitally monitor almost all space in different ways. We have

[40] https://www.asc.upenn.edu/news-events/publications/tradeoff-fallacy-how-marketers-are-misrepresenting-american-consumers-and

already seen the emergence of coded spaces[41] in almost every area of modern life, but it is fair to assume that in the future privacy settings will also become a characteristic of space, and will influence the behaviour of people within it. Urban spaces will exist on a continuum of more- to less-surveilled, with their position on that continuum influencing the way people use them. In fact, as this report shows, people are already aware of influencing space by being monitored in various ways. Research bridging the previously unlinked fields of data ethics and urban geography may help cities understand how space is being produced by citizens, and how choice can be exercised on both sides to negotiate a city that is open to everyone.

Ultimately the role of the city authorities in the future will be to balance two sometimes contradictory rights: people's right to visibility, action and representation, and their right to autonomy and privacy. Traditionally city governance has not involved thinking about the second: the role of authorities was mainly to gather data and then use it to conceptualise and execute policy. The city of the future, however, will involve a balancing act between the two imperatives, and authorities will have to evolve ways of influencing, nuancing and regulating visibility in urban space. This is an entirely new task, and one that will require new forms of public consultation, rulemaking and enforcement. Smart city information infrastructures are in a state of emergence: it is up to those in charge to ensure that checks and balances evolve in parallel with them if the city of the future is to be not only efficient and safe, but also human and liveable. Only city governments themselves can determine whether people in the smart city will be customers, users or citizens.

5. Annex: Interviews and focus groups

Expert Interviews

Name	**Institution**
Evert Meijer	Geodan
Peter van der Mede	DAT.Mobility
Tom Demeyer	De Waag Societeit
Dorien Zandbergen	Sociologist, University of Amsterdam
Bart van der Sloot	Institute for Information Law, University of Amsterdam
Luca Bertolini	Professor of Urban Planning, University of Amsterdam
Marten den Uyl	CEO, Sentient
Ger Baron	CTO, Amsterdam
Serge Hoogendoorn	Professor, TU Delft
Hans van Lint	Professor, TU Delft
Hiddo Huitzing	PBL Netherlands Environmental Assessment Agency
Matthijs Kouw	PBL Netherlands Environmental Assessment Agency
Sander Klous	Professor of big data ecosystems, University of Amsterdam; KPMG
Berent Daan	Chief Data Scientist - Director of Research, Information and Statistics in Amsterdam

[41] Kitchin, R., & Dodge, M. (2011). *Code/space: Software and everyday life*. Mit Press.

Jan Holvast	Historian and privacy researcher
Luuc Posthumus	Chair, Amsterdam committee for the protection of personal information (CPA)
Bart de Groot	Head of Beehive, ZZP flexwerk space
Mariska Majoor	Director, Prostitution Information Centre Amsterdam
Rence Damming	Privacy Officer, KPN
Arnan Oberski	Issue Manager (lighting), Amsterdam city government

Focus Groups

Focus Group	Gender	Age	Characteristics
People at high risk of being profiled	5 men	20-25	Mixed ethnic and religious backgrounds, students.
Non-users of smart phones	3 men, 4 women	20-65	EU origin: Dutch, Italian, German, Belgian. Different professional backgrounds
Sex workers	5 women (conducted sequentially, 3 + 2)	23-45	Mixed ethnic and national origin, different types of sex work (online/offline)
Non-EU immigrants	1 woman, 6 men	27-68	Spanish and Portuguese native speakers from Europe and Latin America; one Ethiopian
EU-immigrants	4 women, 3 men	25-45	Different professional backgrounds, moved to Amsterdam from non-Dutch EU countries
Freelancers (ZZP'ers)	5 men, 1 woman	40 - 55	Working in information technology development and urban design, two visits during social gatherings of freelancers, one group interview
Technology developers	4 men, 1 woman	25-40	Mixed ethnic and professional backgrounds, all Dutch
Children	3 boys, 3 girls	15-17	Mixed ethic and religious backgrounds, high school students.