Chapter 11: Can Big Data Make a Difference for Urban Management?[1]

Ulrich Mans, Sarah Giest, and Thomas Baar

11.1 Introduction

The term "big data" has emerged as a powerful technology trend affecting many aspects of life. Since the early days of big data applications in science and various commercial sectors, the term has come to refer to the exponential increase in the volume and variety of data available, as well as the availability of new tools and approaches to process ever more complex data. Reflecting its global impact on societies, the United Nations speaks of a "Data Revolution" (UN IAEG 2014). Within several domains, big data are already being applied with success. The increased availability of consumer data, for example, provides new opportunities for business and commercial enterprises to develop targeted advertisements and increase revenues (Mayer-Schönberger and Cukier 2013). Big data have facilitated major scientific breakthroughs in various academic disciplines including healthcare, environmental studies, and physics (Krumholz 2014; Bryant et al. 2008). In the public policy realm, the collection and processing of personal data has already transformed intelligence and surveillance practices (Lyon 2014). Law enforcement is another field that has experienced a growing number of experiments in data-driven innovations, such as fraud detection, crime fighting, and violence (Technopolis et al. 2015).

Given the above-average connectivity in urban areas, cities lie at the heart of the trend towards data-driven approaches for confronting societal challenges (Barber 2013; Thakuriah et al. 2015). With more than half of the world's population residing in cities and more than 90 percent of the population growth through 2050 expected to occur in urban areas, there is increased pressure to look for data-driven solutions in the urban context (Pfeffer et al. 2015). This holds particularly true for cities in the Global South, where urban sprawl represents a

[1] An earlier version of this chapter was presented at the International Studies Association Annual Conference in March 2016, in Atlanta, Georgia, United States.

https://doi.org/10.1017/9781316647554.013 Published online by Cambridge University Press

major impediment to sustainable development. Since the 1970s, low-income cities have experienced a 325 percent population increase. In Latin America alone, 110 million people out of 558 million urbanites live in slums, or so-called no-go areas, where basic municipal utility and service delivery remain scarce (de Boer 2015; Muggah 2015; see also Chapters 7, 8, and 9). In this context, recent studies emphasize that "cities ... are unable to respond to the needs of their growing populations faced with rising violence, crime, and poverty" (Mancini and Súilleabháin 2016: III). Urban scholars argue that many cities are set to struggle with income and social inequality; youth unemployment; homicide and criminal violence; poor access to key services; high concentrations of, or preexisting, violence; and exposure to environmental threats (Muggah and Diniz 2013).

To date, most big data applications in the urban context have centered on the quick wins of managerial practices. For example, data analytics are being used in a variety of urban policy sectors, such as public health or infrastructure improvements. These schemes are often driven by cost-saving considerations (Batty 2013), while there is much less movement vis-à-vis the underlying dynamics of urban life and policies aimed at improving social cohesion. Applications are also mostly occurring in OECD countries, where data generation to date is still much more meaningful than in data-poor regions: Using mobile phone records to improve public transport, for example, is only viable once a certain threshold of mobile phone users and representation across the population has been reached. Such an effort makes sense in affluent cities, but not (yet) in urban agglomerations where the digital infrastructure and connectivity are more nascent. At the same time, there is an increasing number of experiments in the developing world, where new data sources are being collected and analyzed for the public good (Bellagio Big Data Workshop Participants 2014).

This chapter aims to contribute to this emerging discourse about how big data can improve urban policy-making, and focuses on the role that this technology can play in building more inclusive cities in the Global South. The authors highlight the need for urban authorities to invest in additional resources as well as meaningful knowledge transfer mechanisms that are in line with the concept of "mobile urbanism." This is particularly important in low-income cities, where policy-makers are driven by the desire to address urban violence and to build more inclusive cities across different constituencies.

11.2 Managing the City in a Digital Age

Data in the urban context can be used in various ways and are applicable to diverse settings. An analysis of 58 initiatives worldwide, performed by Technopolis, the Oxford Internet Institute, and Centre for European Policy

Studies in 2015, shows that the most widespread use of data relates to agenda setting and/or problem analysis. The same study found that open data were commonly used for transparency, accountability, and increasing participation, whereas administrative and statistical data were used for implementation and monitoring purposes (Technopolis et al. 2015). To understand these applications, we clustered them into three dimensions: *data, processes*, and *community*.

11.2.1 Dimensions of Big Data in the Urban Context

First, big data are about the availability of *data* as a source of information and, ultimately, knowledge. The proliferation of information and communication technologies has led to a data surge. Datasets have become so large and complex that traditional tools and approaches are often inadequate for processing them. While the volume of data that is becoming available is an issue, three additional challenging characteristics of the new complexities of digital data streams are velocity (speed of data streams); variety (unstructured versus structured data streams); and veracity (quality of data) (Soubra 2012). Some have added a number of other Vs, *such* as *viability*, for contexts in which reliable data collection is extremely difficult (Mans and Baar 2014).

Second, big data relate to the development of new tools and practices in order to collect, analyze, and work with this digital information (Mayer-Schönberger and Cukier 2013). King (2013) argues that big data are about the *processes* through which we can generate knowledge. Challenges include capturing, verifying, cleaning, storing, sharing, searching, analyzing, visualizing, and presenting the data. In order to infer information and knowledge from data, new disciplines and practices have started to emerge. Such data sciences are producing highly automated approaches, such as machine learning and pattern recognition. In many instances, however, the interpretation of data is unlikely to be taken over by automatic processes; there are growing concerns about the limitations to technically mediated solutions (see, for example, Latonero et al. 2017). Instead, there is a need for hybrid sets of skills that combine human and machine intelligence for supporting policy decisions.

Third, the growing interest in big data has created a new *community* around digital pioneers, which represents a paradigmatic shift in how a diverse set of stakeholders interacts (Letouzé et al. 2015). In a hyper-connected world, the design and implementation of data-driven innovations are incredibly complex and lead to a shift of existing power balances: data sources are becoming more decentralized and analytical tools more accessible to the wider public. As a result, there are limits to the level of "control" that public authorities have over what happens within local policy networks. At the national level, we already see a myriad of citizen networks starting to engage in decision-making processes through data-driven innovations.[2] We also observe a growing number of professionals in the public policy domain that are warming up to the possibilities that data can bring for improving service delivery to citizens (see Chapter 10). In other words, policy-making in a digital age calls for a more active involvement of new (often loosely connected) stakeholders – such as civil society, private enterprises, or private citizens that hold or produce relevant data (WEF 2015a) – which are able to collect, process, validate, and interpret these newly available types of data.

Big data should therefore be understood as a phenomenon bringing together a large variety of stakeholders that individually or collectively engage in the processes that determine how data are collected and used for, among other things, policy goals. Here, it is important to differentiate between data-driven and data-informed policy. Rather than relying on data alone, the term "data-informed policy" refers to decisions that include data as just one factor, coupled with more qualitative judgments about context and potential risks.

The following section presents the academic discourse on knowledge management in cities that applies in the context of data-driven innovation. The subsequent sections look at the different data types that shape the Data Revolution landscape and reflect on their potential benefits. We base this reflection on two case studies that highlight the intricacies of knowledge transfer for effective integration of data-driven innovation into urban policy development: data-informed policy.

11.2.2 Addressing the Urban Knowledge Gap

With the emergence of a large variety of data streams that offer (real time) information on what happens in the city, urban authorities around the world have started to explore new opportunities for improving traffic oversight, service delivery, or crime fighting. At the same time, there are limitations to data-driven innovation. Major barriers are the lack of capacity to apply the insights derived from big data and the inability to effectively inform decision-making using big data in specific cases. To date, many local governments are not equipped for using big data; therefore, capacity-building is considered a pressing challenge (van Edwijk et al. 2015; Giest 2017).

Recent literature offers various models for gaining knowledge on urban dynamics, and how to operationalize these for improved and better-informed decision-making. On the one hand, knowledge management is discussed as

[2] Examples include the Kenyan citizen engagement platform, *Ushahidi* (see: https://www.ushahidi.com/), or Latin American initiatives such as *Chequeado* (see: http://chequeado.com).

https://doi.org/10.1017/9781316647554.013 Published online by Cambridge University Press

a city-specific issue; on the other, there is a discourse on knowledge transfer between cities. Both play a crucial role in understanding the dynamics of data use for urban policy-making.

The Learning City 1: Policy Transfer versus Mobile Urbanism

For city-to-city knowledge transfer, there are two slightly different conceptual models of how knowledge is transferred. First, there is the political science understanding of "policy transfer," which describes an unstructured market of policy ideas that are adopted, transferred, or emulated to maximize reform goals (Peck and Theodore 2010). Put differently, policy transfer is a process in which "knowledge about how policies, administrative arrangements, institutions and ideas in one political setting (past or present) is used in the development of policies, administrative arrangements, institutions and ideas in another political setting" (Dolowitz and Marsh, 2000: 5). The idea of policy transfer has increasingly been paired with the concept of learning in order to understand better how the information that is being transferred is shaped and used in the local context. This, in turn, has led to a discussion about different forms of learning, depending on the political pressure on, as well as the capacity of, policy-makers to adopt new ideas (Giest 2016). Cohen and Levinthal (1990) highlight that "learning capabilities involve the development of the capacity to assimilate existing knowledge" (quoted in Giest 2016: 130). Learning also plays a role in related policy transfer models, such as Municipal International Cooperation (MIC) and city twinning. These are collaboration schemes among two or more cities aiming to transfer knowledge based on a formal relationship. By definition, MIC takes the form of a collaborative effort between local governments to stimulate knowledge exchange between their staff members, often on previously identified topics (van Edwijk et al. 2015). MIC tends to serve broader political goals, such as strengthening democracy and enabling city diplomacy relations, than city twinning. The idea of city twinning builds on a similar idea. Here, cities in distinct geographical and political areas are paired, mainly between North American or European cities and African or South American cities (Muggah 2014).

Next to policy transfer, there is a more recent approach referred to as "policy mobility" or "mobile urbanism." This approach highlights the translational, networked, and multiscalar nature of urban policy (McCann and Ward 2011). The main difference vis-à-vis policy transfer is that mobile urbanism includes a broader set of actors, going beyond policy-makers and bureaucrats to include players who can come from anywhere inside or outside the city. Examples include local policy-makers who use best practice cases from other places and global communities that are adapted to the local context. Here, practitioners

https://doi.org/10.1017/9781316647554.013 Published online by Cambridge University Press

emphasize the need to balance local impacts on the one hand and global flows of knowledge on the other (Dicken et al. 2001; McCann and Ward 2010).

When discussing urban policies in a digital age, the high degree of "mobility" of ideas is particularly relevant to data-driven innovation. Technology advances are fast paced, and if innovative solutions in a given city have proven successful, these can travel quickly to inspire policy-makers in other cities that face similar challenges. At the same time, this knowledge/policy transfer is often a highly political one, as there are struggles related to which policies are being framed as successes, thus empowering certain cities at the expense of others (Robinson 2006; McCann and Ward 2010).

The Learning City 2: Knowledge Management within Cities

Before policies can travel between cities, the research and practice communities within a city play a crucial role in developing successful measures when it comes to introducing new routines and innovative practices (Mans and Meerow 2012). For big data applications, in particular, policy-makers are largely dependent on external advice and input from scientific institutions, technology companies, or related sets of experts to inform or guide decision-making. Knowledge or information management can thereby take various forms. In the urban context, researchers highlight the role of local citizens and their participatory role in the process of developing localized types of knowledge (Hordijk and Baud 2006; Mancini and Súilleabháin 2016). With respect to big data applications in policy development, local governments have often relied on data collected by other actors in the city, or even at the national level. "The result," they note, "is a highly fragmented and dispersed set of local level data" (Hordijk and Baud 2006: 675). In addition, local knowledge is crucial for understanding how to account for biases in big data (that is, representativeness of the local community) and how to provide the required context for analysis (Taylor 2015). These necessities lead to an emphasis on building networks that connect the relevant stakeholders to enable a more critical reflection and improved understanding of the data, informed by local and contextual knowledge. As a report by the Aspen Institute (2012: 11) points out,

> [The integration of data-driven innovation in policy development] will require training a cadre of individuals and intermediary organizations to understand neighborhoods as well as statistics and using "data coaches" to community groups. To be effective data coaches, individuals and organizations must be responsive to communities and their priorities, get better at "translation work" that allows them to interpret data and present it in forms that are useful to practitioners, and develop tools and strategies that make it easier for practitioners to use data for self-evaluation and decision-making.

It is not enough to develop an infrastructure for transferring information and data. Cities need to invest proactively in a strategy that connects citizens and policy-makers to foster data-driven innovation. City authorities need to put in place a new type of digital communications environment and adequate mechanisms when integrating data-driven innovation as part of their operations and policy-making. Such changes can take the form of individuals, institutions, and/or technologies, as well as through importing models from other cities (Komninos 2002; Fuggetta 2012). In this process, it is important to account for the speed of innovation in data-related technology: it is increasingly difficult to keep a sufficiently up-to-date overview of all relevant developments, even if there are enough resources for a dedicated team of experts. Instead, city authorities increasingly have to rely on hybrid, international networks of experts that share best practices as these emerge from pilot projects around the globe (Verhulst 2016).

11.3 Towards More Inclusive Cities? Tackling Inequality and Violence with Data

How can big data help policy-makers build more inclusive cities in the Global South? There are many ways to approach this question; for the purposes of this chapter, we focus on the possibilities that are emerging for tackling inequality and violence. We first present five categories of data streams, and then present the possible impact these could have on both challenges. Even though using big data to accomplish inclusivity goals is a relatively nascent field, we present some insights from published case studies on reducing violence in cities within Colombia and South Africa to highlight recent developments in the use of data and the knowledge transfer mechanisms involved.

11.3.1 (New) Types of Data Streams

When looking at the opportunities and challenges that come with the Data Revolution, it is useful to distinguish between various categories of additions to the data landscape that have entered (or are likely to enter) the city's policy realm. It is important to note that much of the big data discourse addresses the emerging possibilities of *data analytics* and new computational methodologies to handle increasingly large databases. For example, technology advances in the fields of real-time dashboards, automated visualizations, machine learning, and artificial intelligence have generated much interest in this regard. However, it is useful to move beyond the analytics, and instead to define the new *types* of data streams that are likely to shape the way decision-making is undertaken.

Many of the more radical, data-driven innovations are inspired by new types of data that have thus far not been collected by city authorities. In this context, Rigobon (2016) refers to "designed" and "organic" data streams, which emphasize that what data will be collected has traditionally been decided beforehand and has subsequently been collected according to a predefined scheme (through surveys, questionnaires, and/or administrative records, for example). The main difference between these traditional data collection regimes and big data collection is that new data streams increasingly come in the form of unstructured data. In the following, we introduce five types of data streams that can help to navigate today's data landscape: public datasets, citizen reporting, open web data, digital breadcrumbs, and remote sensing.

Public Datasets

Although public datasets do not necessarily constitute a new type of data stream, digitization and the availability of new analytical capacities lead to an increased uptake of these data in policy-making processes. Data sources for policymaking now include, a.o. "real-time sensor data, public administration data (including open data), data from statistical offices, commercially traded data and several types of targeted or ad-hoc data" (Technopolis et al. 2015: n.p.). In addition, we observe the promotion of open data in the public sector and among NGOs, which leads to increased free availability of these datasets in machine-readable formats. The digital divide is still a major limiting factor in this form of data collection. Governments in non-OECD countries are generally much more reluctant – and less able – to make datasets publicly available.[3] Questions remain regarding the extent to which digital technologies can improve the collection of data in the developing world, and how much of this additional data will be made available for urban authorities (or other third parties) as a consequence.[4]

Citizen Reporting

With access to mobile devices and the Internet on the rise, connecting to citizens is becoming cheaper, faster, and more reliable. This connectivity can be used for survey techniques based on Short Message Service (SMS), online feedback forms, and so forth. Collecting data in this way is often conducted through digital platforms, which can be run by public entities, private or

[3] As part of its Global Open Data Index, Open Knowledge International provides an overview and comparative ranking on open government data (OKI 2014).

[4] In January 2017, the first UN World Data Forum took place in Cape Town, South Africa. At the meeting, national statistics officials and data and technology experts held numerous meetings to discuss how to apply new data technologies to monitor progress on the Sustainable Development Goals.

community organizations, or as a joint effort. In various Kenyan cities for example, the NGO Sisi ni Amani applied SMS-based citizen reporting in order to reduce ethnic tensions across communities (Parker 2011; Trujillo et al. 2013); other examples include violence monitoring at several protest sites in Bangkok throughout 2014, "in order to better understand the situation and track relevant developments" (Elva 2014: n.p.). Further, the Nairobi police have been experimenting with the use of cell phones to reach out to slum inhabitants in Mathare (Frilander et al. 2014). Even in such underserved areas of the city, mobile phone ownership is nearly universal, and approximately 50 percent of these devices are Internet enabled, which makes direct, real-time communication with citizens a possibility (whether by police or other public services agencies). Still, particular challenges can arise with regard to the validity and representativeness of the information provided by respondents in this style of big data collection (van der Windt, 2012).

Open Web Data

Online content has long been readily available in the form of websites, news archives, event reporting, and blog posts. This includes online platforms such as Global Dataset of Events, Language, and Tone (GDELT) or Armed Conflict Location & Event Data Project (ACLED) that provide event data,[5] or simply search engine tools that are available to any online reader.[6] New developments include a) an increasing number of methodologies making it possible to "scrape" the content of websites automatically without human oversight and b) the emergence of social media as an additional form of open web data. Popular platforms including Facebook, Twitter, Instagram, YouTube, and LinkedIn, as well as many other social platforms, offer various degrees of access to their customers' data.

To be clear, the latter is a peculiar form of "open" data. Many of these sources are available to the general public, yet access to them is controlled by private entities. Depending on the aims and privacy restrictions that come with the use of this type of data stream, it is possible to derive relevant insights from what is posted online. These insights can be used for assessments of political preferences and social topics of interest extrapolated from Twitter messages (UN Global Pulse 2014), to verify flood damage across urban settlements using multiple social media platforms (Quaggiotto 2014), or to analyze social patterns in relation to security/crime issues in the context of cities (Pfeffer et al. 2015). It is also to possible establish knowledge of social and political networks

[5] www.gdelt.org and www.acleddata.com

[6] See, for example, www.forbes.com/sites/kalevleetaru/2015/09/28/is-the-black-lives-matters-movement-fading-a-data-driven-look-at-web-searches-and-television/.

based on this data (O'Callaghan et al. 2014; Bozdag et al. 2014). It is likely that many of today's possibilities will evolve in the coming years. The key question is which open online data streams can be employed to gain relevant insights for users, and to what extent machine-readable is access granted?

Digital Breadcrumbs

The more people are connected to or work with digital technologies, the more they leave traces of what they do in their daily lives (Pentland 2012). This includes any type of consumption in digital form (supermarket purchases, cell phone airtime vouchers, or financial transfers). Even though this type of data is not necessarily representative, it can reach far beyond the middle class. For example, refugees receive vouchers in the form of e-cards that register what, when, and where people buy goods (WFP 2017; Flaemig et al. 2017). To date, the most powerful form of these "breadcrumbs" are mobile phone data. There are a number of interesting experiments with cell phone data, for example, to detect crime hotspots in London (Bogomolov et al. 2015) and understanding social ties across different communities in the Ivory Coast (Bucicovschi et al. 2013). Also, mobile phone data have been used in Afghanistan to determine changes in movement patterns after micro-violence, such as improvised explosive device (IED) explosions (World Bank 2014), and to develop new poverty monitoring methodologies in Senegal (Pokhriyal et al. 2015). However, digital breadcrumbs come with major caveats.

On the one hand, these types of data streams are often proprietary and not accessible without prior negotiations with a commercial party, such as telecom providers or financial service providers. Second, the clients of these services do not generally know about (or consent to) their data being used (this is different, for example, than social media content, for which a certain degree of consent can be assumed). Even though analysis of digital breadcrumbs is generally done on an aggregated level without substantial risks of privacy infringements, full privacy does not exist: Most datasets that include personal data carry the risk that individuals can be reidentified (Berens et al. 2016; OCHA 2016). Currently, standards for data sharing and data use simply do not exist to a degree that makes all stakeholders comfortable with experimentation with these types of datasets. However, sector-specific data use guidelines and related frameworks that help create trust and form new data collaboratives are likely to emerge over time (WEF 2015b; IDRG 2015; GovLab 2016).

Remote Sensing Data

Satellite images are a well-known source of data that are usually expensive, but are increasingly accessible, even for smaller organizations. This technology is based on sensors that have been placed in orbit, made possible only via monetary investments. The affordability of remote sensing has risen in part because common sensors are being placed nearly everywhere, from closed circuit television cameras to air quality sensors, track-and-trace devices in vehicles, and sensors required for the Internet of Things (for example, sensors in refrigerators, street lights, and so forth). An interesting example is the ability, through remote sensing, to "measure the quantity, timing, and locations of gunfire incidents with greater accuracy than do reported crime or 911 call data through sensors" (Carr and Doleac 2016: 4). This technology, called "Shotspotter," is currently applied in the United States (ibid.). Shotspotter's physical manifestation is a connected system of audio sensors on top of buildings that detects the sounds of gunfire and analyzes them for accuracy. If Shotspotter confirms the sound of gunfire, the program responds by sending a message to local police with the location of the shots fired. The data produced by Shotspotter – date, time, location, single/multiple gunshots – are publicly available.

Likewise, in the geospatial arena, the emergence of drones as a new type of cheap sensor increasingly impacts the way environmental data can be collected or verified. In disaster areas, for example, drones are already being used for quick damage assessments, and a growing number of experiments are underway to use drone-mounted cameras in the fields of agriculture or environmental protection in urban areas (see, for example, Meier 2014). Affordable, high-resolution satellite imagery enables people to retrieve information about hard-to-reach places and conflict areas. For example, "Amnesty International requested the assistance of the Geospatial Technologies and Human Rights Project of the American Association for the Advancement of Science to investigate the veracity of reports of human rights violations stemming from the escalating conflict in Aleppo, Syria" (Amnesty International n.d.: n.p.).

These five types of data streams can have different applications in different contexts. Looking at the innovation landscape today, we see a number of cases that address aspects of urban violence, that is, policing, law and order, and related challenges. Examples of more structural approaches that use data-driven innovations to reduce inequality throughout the city are less common.[7] This is not a surprise, as many questions remain about the extent to which new data streams can complement classical data sources, especially in a developing country. Data are generally biased towards the digital haves and have-nots; we need to develop methodologies that make new data streams both representative and reliable. Table 11.1 gives an overview of the possible uses of these five new types of data streams for both the reduction of violence and inequality in urban contexts.

[7] Exceptions include http://masschallenge.org/startups/2016/profile/ubuntucapital.

Table 11.1 Possible uses for data in creating more inclusive cities

	Examples of data application	
Type of data stream	Reducing violence	Reducing inequality
Public datasets (Census and administrative data on policing, education, healthcare, and so forth)	Data from police reports can be matched with other data streams such as SMS-based surveys.	Census data can be used in combination with social media content to understand public perceptions among youth, for example, on unemployment.
Citizen reporting (SMS-based surveys, online reporting platforms, and so forth)	Police departments can collect information from citizens on crime-related incidents in a given area.	Local perceptions of major issues in a given area can be collected by public authorities and/or local community-based organizations.
Open web data (Online content, social media, and so forth)	Social media can be used to identify hate speech towards a given group; it can also be used for outreach purposes to encourage citizens to avoid certain areas or not to engage in violence.	Social media content can be collected and analyzed in order to determine major problems in certain areas or to encourage civic engagement.
Digital breadcrumbs (Consumer data, mobile phone data, and so forth)	Aggregated mobile phone data can show where people move at night, giving clues about relative safety in certain urban areas.	Aggregated consumer data (for example, airtime vouchers) can reveal major changes in the socioeconomic situation of certain areas.

https://doi.org/10.1017/9781316647554.013 Published online by Cambridge University Press

Remote sensing		
(Satellite imagery, sensor networks, Internet of Things, and so forth)	Audio sensors can detect gunshots in real time and provide clues about the deterioration of security in a given area.	Air- or water-quality sensors can detect problems with the quality of public goods.

As discussed in the previous section, any of these applications requires a meaningful dialogue between those who work with the technology and those with contextual expertise regarding the location in which it will be applied. We are at the very beginning of the Data Revolution – much remains unexplored and untested; indeed, the use of new data streams in formulating city policies is far from mainstream. City authorities tend to start with existing data rather than tapping into new data streams. Moving forward, we need to improve our understanding of the underlying dynamics of knowledge transfers insofar as they relate to data-driven innovation. While still evolving, two examples, from Cali and Cape Town, highlight some of the lessons learned about knowledge transfer mechanisms that support data-informed policy.

11.3.2 Reducing Violence with Data Knowledge: Cali and Cape Town

Cali – Colombia, and Cape Town – South Africa are two cities that have shifted towards data-informed policy in connection to reducing violence. We identify some of the opportunities and challenges that are connected to this shift. Generally speaking, the availability of additional data has led some cities to take a more evidence- and/or data-based approach towards violence; Colombia has become an especially popular research example (see Gaviria 2000; Bourguignon et al. 2002; Cotte Poveda 2012).

In Latin America, several cities – including Bogotá, Cali, Medellín, San Pablo, and Recife – have been able to reduce violent incidents dramatically using policies that harness big data. The programs stem from a mixture of models used in the United States and evidence for what works in the targeted cities in Latin America (Ojea 2014). This has also led to new revelations about the root causes of violence. For example, for a long time, the US lens on crime, in combination with substantial media coverage of drug-related crimes, led officials in Cali to believe that drug dealers were the biggest cause of homicides in the city

(Velasco 2015). Using recent and local statistics, however, officials learned that "homicide victims and aggressors were predominantly young, unemployed males who had low levels of education, came from the poorer sectors of the city and were frequently involved in gang fights" (Velasco 2015: 3). In other words, drug traffic was still part of the equation, but was only indirectly responsible for violence. The crime figures in this case largely came from an online platform called "The Monitor," which interactively maps the distribution of murder by country, year, age of victim and, where available, gender, and type of weapon. The online database draws on statistics from the United Nations Office for Drugs and Crime, government offices, health institutes, and policy records, as well as a detailed, city-level breakdown for Latin America. However, streamlining such information is challenging, since Latin American countries have different ways of defining crime and differ in the way they collect information. The Inter-American Development Bank is currently in the process of standardizing violence indicators (Velasco 2015).

Cape Town has also moved towards a more comprehensive approach for tackling violence based on quantitative and qualitative data. This shift was facilitated by the Violence Prevention through Urban Upgrading (VPUU) not-for-profit initiative, which works with local and national governments and includes international groups with stakeholder expertise in developing such measures. The VPUU applied a combination of high-quality, research-based documentations, monitoring, and evaluation surveys, as well as databases of police-reported robberies over a ten-year period (Cassidy et al. 2015), as well as incorporating census data and information from the South African Index of Multiple Deprivation. The researchers subsequently geolocated the data to specific areas through the use of mobile phones that were distributed to the community (Cassidy et al. 2015). In this way, citizen reporting, digital breadcrumbs, secondary databases, and qualitative information were gathered to inform potential policy changes. These changes have led officials to focus increasingly on infrastructural causes for violence, such as lighting, improved public spaces, and safer public transportation, after-school activities, and an improved education system (WCG 2011; Cassidy et al. 2015).

In both cities, a diverse set of stakeholders initiated policy changes to incorporate big data. Cali's mayor, Dr. Rodrigo Guerrero, introduced weekly meetings of the heads of all departments connected with law enforcement (Rosenberg 2014). Those meetings involved officials from "the police, judiciary and forensic authorities, members of the Institute for Research and Development in Violence Prevention and Promotion of Social Coexistence (CISALVA) at the University of Valle, cabinet members responsible for public safety, and the municipal statistics agency" (Velasco 2015: 6). The meetings were an attempt to pool contextual knowledge on violence in combination with the data to make sense of the status quo and to discover possible improvements to initiatives. In Cape Town, as in Cali, the goal was a more comprehensive approach to violence. Here, changes involved the inclusion of stakeholders in public health, criminal justice, education, and social development sectors, and active participation and partnership of citizens and civil society more broadly (WCG 2013; Cassidy et al. 2015).

Both cities also faced political obstacles, including changes in local government, funding, and knowledge sharing among local stakeholders. For Cali, these challenges were twofold: first, the national government was unwilling to provide additional financial support to data-driven innovation. The city needed money to support more policing in risk-prone areas, during holidays and paydays, as well as after 2 a.m. – days and times during which violence had been shown to increase. In addition, because Colombian mayors can serve only one term, newly implemented measures could be, and were, overturned by the new mayor. After Mayor Guerrero's term (1992–1995), the murder rate rose again (Rosenberg 2014). In Cape Town, measures suggested by VPUU were unpopular with the government because they targeted areas where the political opposition was in charge. According to Cassidy et al. (2015), this not only resulted in limited implementation, it further posed a threat to the research process, since it compromised the availability and validity of evaluative data from community stakeholders and drove an overreliance on administrative data. Ultimately, crime data can also be uncomfortable for mayors and governments, especially before elections, since better recording and more accurate data often lead to higher reported crime rates that might hurt political ambitions.

Overall, both cities are increasingly incorporating data-informed policies into their measures against violence and have, over the course of establishing these initiatives, involved a range of stakeholders who can provide more contextual perspectives. In the years to come, additional data tools could lead to more accurate and complete data on crime and violence trends in cities. However, as the examples have also shown, there is a political component that can slow down or even hinder the use of big data.

11.3.3 Discussion

Our examples from South Africa and Colombia show that data-informed policy is largely shaped through joint efforts of national and local governments as well as local communities and law enforcement agencies. These case studies also indicate that data are only one piece of the larger puzzle when targeting violence in cities; issues remain surrounding political and collaborative aspects. To guide future paths for data use in the context of urban policy in the Global South, we believe there are two overarching lessons summarized by these cases.

Data-Informed Policy-Making

First, using big data is accompanied by risks of drawing misleading conclusions, such as assumptions about causes of violence that are drawn from public datasets, but do not apply to a specific region. Data analytics cannot simulate the complex picture of potential interactions of different policy domains, such as crime and infrastructure, or the dynamics among social groups in certain neighborhoods (Bollier 2010). The research community is skeptical of claims of universal urban experiences, stressing that contextual particularities and local experiences within places are important (Brenner and Schmid 2015; Thakuriah et al. 2015). It follows that conclusions drawn in cities with high crime rates do not automatically apply to other cities with similar statistics, but different local contexts. The example of Cali has shown that officials were too quick to assume that drug-related crime was driving up the homicide numbers when drug trafficking had only an indirect effect. However, the challenge is to strike an appropriate balance between automated analysis and contextual interpretation now that data are becoming more widely used.

The Politics of Data-informed Policy

Second, data can be political. When utilizing the information gained from data, political obstacles emerge in two ways. Data can bring to the surface insights that are uncomfortable to political stakeholders. Cape Town exemplifies a city uncooperative in data collection efforts, either because proposed data collection efforts were connected to regions in the hands of the political opposition or because data collection initiatives were branded as campaigns against the government (Consortium on Crime and Violence Prevention 2015). Furthermore, collaboration across political constituencies might prove difficult. Based on the insights from Cali and Cape Town, cross-stakeholder engagement emerges as a key dimension for deploying data-based initiatives in cities. Such engagement has been achieved in the form of regular meetings of heads of departments (Cali) or by involving citizens in data collection (Cape Town). Underlying this collaboration is the notion of trust – trusting that the data are put to good use by government, as well as trust in local stakeholders by the government. Moving towards more data-informed policies, city stakeholders will have to find meaningful ways to create mutual trust.

The elements discussed in this chapter call for a more thorough understanding of how advances in data-driven innovation could translate into new forms of urban policy-making – and how collaboration between various stakeholders and actors can be supported from the beginning to avoid inappropriate technology and policy designs. Much remains to be done to support decisions about which policies to adopt and when to be cautious in applying data-informed policy. From a research perspective, future studies should give clues about the interplay of additional, more detailed data being collected and the political repercussions this might have. If new data streams enable more accurate, but also more problematic, numbers for certain issues such as violence and poverty, the political opposition might outweigh the societal benefits that data-driven innovations provide. Overcoming these obstacles requires alignment between different stakeholders within the city, as well as paying attention to the timing and circumstances within which data-informed policies are developed.

References

Amnesty International (n.d.) *Satellite Imagery Analysis for Urban Conflict Documentation: Aleppo, Syria*, AAAS. www.aaas.org/aleppo.

Aspen Institute (2012) Developing and Using Data and Evidence to Improve Place-Based Work. Proceedings from a meeting convened by the Aspen Institute Roundtable on Community Change with support from The Annie E. Casey Foundation, New York, September 8–9, 2011. https://assets.aspeninstitute.org/content/uploads/files/content/images/rcc/Sept%202011%20Data%20Meeting%20Proceedings%20FINAL%206%206%2012.pdf

Barber, B. (2013) *If Mayors Ruled the World: Dysfunctional Nations, Rising Cities*. New Haven, CT: Yale University Press.

Batty, M. (2013) Big Data, Smart Cities and City Planning. *Dialogues in Human Geography*, 3(3): 274–279.

Bellagio Big Data Workshop Participants (2014) *Big Data and Positive Social Change in the Developing World: A White Paper for Practitioners and Researchers*, Oxford: Oxford Internet Institute. www.rockefellerfoundation.org/uploads/files/c220f1f3-2e9a-4fc6-be6 c-45d42849b897-big-data-and.pdf.

Berens, J.B., Mans, U., and Verhulst, S. (2016) Comparing and Mapping Responsible Data Approaches. TheGovLab and Center for Innovation, Leiden University.

de Boer, J. (2015) Resilience and the Fragile City. *Stability: International Journal of Security and Development*. 4(1): p.Art. 17. doi: http://DOI.org/10.5334/sta.fk.

Bogomolov, A. Lepri, B., Staiano, J., Letouze E., Oliver, N., Pianesi, F., and Pentland, A. (2015) Moves on the Street: Classifying Crime Hotspots Using Aggregated Anonymized Data on People Dynamics. *Big Data*, 3(3): 148–158.

Bollier, D. (2010) The promise and peril of Big Data, Aspen Institute, available online: www.emc.com/collateral/analyst-reports/10334-ar-promise-peril-of-big-data.pdf.

Bourguignon, F., Nunes, J., and Sanchez, F. (2002) What Part of the Income Distribution Matter for Explaining Crime? The case of Colombia. Working Paper, DELTA ENS.

Bozdag, E., Gao, Q., Houben, G.J., and Warnier, M. (2014) Does Offline Political Segregation Affect the Filter Bubble? An Empirical Analysis of Information Diversity for Dutch and Turkish Twitter Users. *Computers in Human Behavior*, 41: 405–415.

Bryant, R., Katz, R.H., and Lazowska, E.D. (2008) Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society. http://cra.org/ccc/wp-content/uploads/sites/2/2015/05/Big_Data.pdf.

Bucicovschi, O., Douglass, R.W., Meyer, D.A., Ram, M., Rideout, D., and Song, D. (2013) Analyzing Social Divisions Using Cell Phone Data, in Blondel, V., de Cordes, N., Decuyper, A., Devlisse, P., Raguenez, J. and Smoreda, Z. (eds.), *Mobile Phone Data for Development*. Cambridge, MA: Netmob, pp. 42–54.

Brenner N. and Schmid C. (2015) *Towards a New Epistemology of the Urban City* 1(2–3): 151–182.

Cohen, W.M. and Levinthal, D.A. (1990) Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1): 128–152.

Carr, J. and Doleac, J. (2016) The Geography, Incidence, and Underreporting of Gun Violence: New Evidence Using ShotSpotter Data. Brookings Research Paper, April 2016.

Cassidy, T., Ntshingwa, M., Galuszka, J., and Matzopoulos, R. (2015) Evaluation of a Cape Town Safety Intervention as a Model for Good Practice: A Partnership between Researchers, Community and Implementing Agency. *Stability: International Journal of Security and Development*. 4(1): p.Art. 27. doi: http://DOI.org/10.5334/sta.fi.

Consortium on Crime and Violence Prevention (2015) Dialogue on Safety and Security: Insights from Cape Town. Key Points from the Third International Dialogue on Safety and Security Cape Town, February 25–27, 2015. https://igarape.org.br/wp-content/uploads/2015/10/SafetySecurityDialoguesReport.pdf.

Cotte Poveda, A. (2012) Violence and Economic Development in Colombian Cities: A Dynamic Panel Data Analysis. *Journal of International Development*, 24: 809–827.

Dicken, P., Kelly, P.F., Olds, K., and Wai-Chung Yeung, H. (2001) Chains and Networks, Territories and Scales: Towards a relational framework for analysing the global economy. *Global Networks*, 1: 89–112. doi:10.1111/1471–0374.00007.

Dolowitz, D., and Marsh D. (2000) Learning From Abroad: The Role of Policy Transfer in Contemporary Policy Making. *Governance*, 13 (1): 5–24.

van Edwijk, E., Baud, I., Bontenbal, M., Hordijk, M., van Lindert, P., Nijenhuis, G., and van Westen, G. (2015) Capacity Development or New Learning Spaces through Municipal International Cooperation: Policy Mobility at Work? *Urban Studies*, 52 (4): 756–774.

Elva Community Engagement [Elva] (2014) Monitoring Violence in Bangkok. https://elva.org/case-studies.

Flaemig, T., Sandstrom, S., Caccavale, O.M., Bauer, J-M., Husain, A., Halma, A., and Poldermans, J. (2017) *Using Big Data to Analyse WFP's Digital Cash Programme in Lebanon*. Humanitarian Practice Network. http://odihpn.org/blog/using-big-data-to-analyse-wfps-digital-cash-programme-in-lebanon.

Frilander, M., Lundine, J., Kutalek, D., and Likaka, L. (2014) *New Technologies for Improving Old Public Security Challenges in Nairobi*. Igarapé Institute. http://issat.dcaf.ch/mkd/content/download/49277/782313/file/NE-15_New-technologies-in-Nairobi_4th_june.pdf.

Fuggetta, A. (2012) 3+1 Challenges for the Future of Universities, *The Journal of Systems and Software*, 85: 2417–2424.

Gaviria, A. (2000) Increasing Returns and the Economic Evolution of Violent Crime: The Case of Colombia. *Journal of Development Economics*, 61: 1–25.

Giest, S. (2016) Overcoming the Failure of 'Silicon Somewheres': Collective Learning in Innovation Networks. *Policy and Politics*. 45(1): 39–54 doi.org/10.1332/030557316X14779412013740.

Giest, S. (2017) Big Data Analytics for Mitigating Carbon Emissions in Smart Cities: Opportunities and Challenges. *European Planning Studies*, 25(6): 941–957 (doi: 10.1080/09654313.2017.1294149).

GovLab (2016) Data Collaboratives: Exchanging Data to Improve People's Lives, GovLab (blog). http://thegovlab.org/datacollaboratives.

Hordijk, M. and I. Baud. (2006) The Role of Research and Knowledge Generation in Collective Action and Urban Governance: How Can Researchers Act as Catalysts? *Habitat International*, 30: 668–689.

International Data Responsibility Group [IDRG] (2015) International Data Responsibility Group: Putting People First in a Digital Age. www.responsible-data.org.

King, G. (2013) Big Data is Not About The Data! Talk at the New England AI Meetup, 5/14/2013. http://gking.harvard.edu/files/gking/files/evbase-neai_0.pdf.

Komninos, N. (2002) *Intelligent Cities: Innovation, Knowledge Systems, and Digital Spaces*. London and New York: Routledge.

Krumholz, H.M. (2014) Big Data and New Knowledge in Medicine: The Thinking, Training, and Tools Needed for a Learning Health System. *Health Affairs*, 33 (7): 1163–1170.

Latonero, M., Kleinman, M., and Hiatt, K. (2017) Tech Folk: 'Move Fast and Break Things' Doesn't Work When Lives Are at Stake, *The Guardian*, February 2017. www.theguardian.com/global-development-professionals-network/2017/feb/02/technology-human-rights.

Letouze, E., Vinck, P., and Kammourieh, L. (2015) The Law, Politics and Ethics of Cell Phone Data Analytics. Data-Pop Alliance publication. http://datapopalliance.org/wp-content/uploads/2015/04/WPS_LawPoliticsEthicsCellPhoneDataAnalytics.pdf.

Lyon, D. (2014) Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique. *Big Data & Society*, July–December: 1–13.

Mans, U., and Baar, T.J. (2014) *Defining Peace Informatics*. www.peaceinformaticslab.org/learn-more.html.

Mans, U. and Meerow, S. (2012) Role of Local Governments in Promoting Renewable Energy Businesses. ICLEI Policy Paper.

Mancini F. and Súilleabháin A.Ó. (eds.) (2016) *Building Resilience in Cities under Stress*. New York: International Peace Institute.

McCann, E. and Ward, K. (2010) Relationality / Territoriality: Toward a Conceptualization of Cities in the World. *Geoforum* 41 (2), 175–184.

McCann, E. and Ward, K. (2011) *Mobile Urbanism*. Minneapolis: University of Minnesota Press.

Meier, P. (2014). *Digital Humanitarians: How Big Data is Changing the Face of Humanitarian Response*. Boca Raton: CRC Press.

Mayer-Schönberger, V. and Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.

Muggah, R. (2014) How Are Some Fragile Cities Tackling Their Worst Problems? [TED Talk]. www.npr.org/2016/01/08/462282897/how-are-some-fragile-cities-tackling-their-worst-problems.

Muggah, R. (2015) A Manifesto for the Fragile City. *Journal of International Affairs*, 68 (2): 19–36.

Muggah, R. and Diniz, G.M. (2013) Digitally Enhanced Violence Prevention in the Americas. *Stability: International Journal of Security and Development*. 2(3): p.Art. 57. DOI: http://doi.org/10.5334/sta.cq.

O'Callaghan, D., Prucha, N., Greene, D., Conway, M., Carthy, J., and Cunningham, P. (2014) Online Social Media in the Syria Conflict: Encompassing the Extremes and the In-Betweens, Proceedings of the 2014 International Conference on Advances in Social Network Analysis and Mining (ASONAM 2014).

OCHA Think Brief [OCHA] (2016) Building Data Responsibility into Humanitarian Action. United Nations Office for the Coordination of Humanitarian Affairs. www.unocha.org/sites/unocha/files/Building%20data%20responsibility%20into%20humanitarian%20action.pdf

Ojea, V. (2014) Latin America Accounts for More than 30% of the World's Homicides. The World Bank, March 5. www.worldbank.org/en/news/feature/2014/02/11/en-america-latina-sufre-mas-del-30-de-los-homicidios-mundiales

Open Knowledge International [OKI] (2014) Tracking the State of Government Open Data. http://index.okfn.org.

Parker, P. (2011) Sisi Ni Amani 'Peacemapping' in Kenya. www.insightonconflict.org/blog/2011/09/kenya-peacemapping.

Peck, J. and Theodore, N. (2010) Mobilizing Policy: Models, Methods, and Mutations. *Geoforum*, 41(2): 169–174.

Pentland, A. (2012) Reinventing Society in the Wake of Big Data. *Edge*, August 30. www.edge.org/conversation/reinventing-society-in-the-wake-of-big-data.

Pokhriyal, N., Dong, W., and Govindaraju, V. (2015) Virtual Networks and Poverty Analysis in Senegal. D4D Netmob, Conference Paper.

Pfeffer, K., Verrest, H. and Poorthuis, A. (2015) Big Data for Better Urban Life? – An Exploratory Study of Critical Urban Issues in Two Caribbean Cities: Paramaribo (Suriname) and Port of Spain (Trinidad and Tobago). *The European Journal of Development Research*, 27(4): 505–522.

Quaggiotto, G. (2014) Using Twitter to Get Ground Truth on Floods: An Interview with Floodtags Founder Jurjen Wagemaker, *UN Global Pulse Blog*, November 19. www.unglobalpulse.org/blog/using-twitter-get-ground-truth-floods-interview-floodtags-founder-jurjen-wagemaker.

Rigobon, R. (2016) Keynote Speech at the Annual Bank Conference on Development Economics 2016: Data and Development Economics, June 20, 2016, Washington, DC.

Robinson, J. (2006) *Ordinary Cities: Between Modernity and Development*. Abingdon and New York: Routledge.

Rosenberg, T. (2014) Colombia's Data-Driven Fight against Crime. *New York Times*. https://opinionator.blogs.nytimes.com/2014/11/20/colombias-data-driven-fight-against-crime/

Soubra, D. (2012) The 3 Vs that Define Big Data. *Data Science Central*, July 5, 2012. www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data.

Taylor, L. (2015) Big Data and Urban Governance, in Gupta, J., Pfeffer, K., Verrest, H., and Ros-Tonen, M. (eds.) *Geographies of Urban Governance*. Cham: Springer International.

Technopolis, Oxford Internet Institute and Centre for European Policy Studies (2015) Data for Policy: A Study of Big Data and Other Innovative Data-Driven Approaches for Evidence-Informed Policymaking. www.technopolis-group.com/data-policy-innovative-data-driven-approaches-evidence-informed-policymaking/.

Thakuriah, P., Tilahun, N., and Zellner, M. (2015) Big Data and Urban Informatics: Innovations and Challenges to Urban Planning and Knowledge Discovery. In Proceedings of NSF Workshop on Big Data and Urban Informatics, 4–32. https://urbanbigdata.uic.edu/proceedings/

Trujillo, H.R., Elam, D., Shapiro, G., and Clayton, M. (2013) The Role of Information and Communication Technology in Preventing Election-Related Violence in Kenya, 2013. *Perspectives on Global Development and Technology*, 13(1–2):, 111–128.

UN Global Pulse (2014) Using Twitter to Understand the Post-2015 Global Conversation, Global Pulse Project Series, no.6, 2014. www.unglobalpulse.org/sites/default/files/UNGP_ProjectSeries_Post2015_Priorities_2014_0.pdf.

UN IAEG (2014) A world That Counts: Mobilizing the Data Revolution for Sustainable Development. www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf.

Van der Windt, P. (2012) From Crowdsourcing to Crowdseeding: The Cutting Edge of Empowerment, in Livingstone, S. and Walter-Drop, G. (eds.). *Bits and Atoms: Information and Communication Technology in Areas of Limited Statehood*, Oxford: Oxford University Press, pp. 144–155.

Velasco, R. (2015) Big Data Are Reducing Homicides in Cities across the Americas. *Scientific American*, October 1, 2015.

Verhulst, S. (2016) Corporate Social Responsibility for a Data Age, *Stanford Social Innovation Review* (February). https://ssir.org/articles/entry/corporate_social_responsibility_for_a_data_age?platform=hootsuite.

Western Cape Government [WCG] (2011) *Western Cape's Draft Strategic Plan*. South Africa: WCG.

Western Cape Government [WCG] (2013), *Integrated Violence Prevention Policy Framework*. South Africa: WCG.

World Bank (2014) Big Data in Action for Development: World Bank Publication 2014. http://live.worldbank.org/sites/default/files/Big%20Data%20for%20Development%20Report_final%20version.pdf.

World Economic Forum [WEF] (2015) *Global Agenda Council on Data-driven development*. WEF Publication (webpage). www.weforum.org/communities/global-agenda-council-on-data-driven-development.

World Economic Forum [WEF] (2015B) Data-Driven Development: Pathways for Progress (January 2015). http://digitalprinciples.org/wp-content/uploads/2015/05/WEFUSA_DataDrivenDevelopment_Report2015.pdf.

World Food Programme [WFP] (2017) Cash-Based Transfer. http://www1.wfp.org/cash-based-transfers.