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)

Pfeffer, K.; Verrest, H.; Poorthuis, A.

DOI
10.1057/ejdr.2015.48

Publication date
2015

Document Version
Final published version

Published in
European Journal of Development Research

Original Article

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)

Karin Pfeffer[a,*], Hebe Verrest[a] and Ate Poorthuis[b]

[a]University of Amsterdam, Amsterdam, Netherlands.
E-mail: k.pfeffer@uva.nl
[b]University of Kentucky, Lexington.

Abstract Big Data is increasingly seen as important in studying the city. This pertains to both its methodological capacity and the societal implications it may have. In this article we draw on contemporary literature to discuss the potentials and challenges of Big Data for addressing pressing urban issues. In addition, we examine the potential of Big Data as a methodological tool for two Caribbean cities, Paramaribo and Port of Spain, for developing new knowledge on urban issues that matter in such cities, specifically water-related risks and security. We do so by interrogating Twitter data to uncover relevant geographical and social patterns of tweets pertaining to water-related risks (Paramaribo) and security/crime issues (Port of Spain) and confronting these with qualitative knowledge about these places. We argue that Big Data are a powerful resource for discovering interesting patterns, but one needs to be critical of the methodological caveats and consider the social-cultural specificities of ICT use.

Les mégadonnées ou 'Big Data' sont considérées comme de plus en plus importantes dans l'étude d'une ville, du fait à la fois de leur capacité méthodologique, mais aussi des implications sociétales qu'elles peuvent avoir. Dans cet article, nous nous appuyons sur la littérature contemporaine pour discuter du potentiel et des défis des mégadonnées pour régler les enjeux urbains pressants. En outre, nous examinons le potentiel de Big Data comme un outil méthodologique pour deux villes des Caraïbes, Paramaribo et Port-d'Espagne, pour développer de nouvelles connaissances sur des questions urbaines primordiales dans ces villes, en particulier sur les risques liés à l'eau et la sécurité. Nous utilisons les données de Twitter pour découvrir des schémas géographiques et sociaux pertinents de Tweets relatifs aux risques liés à l'eau (Paramaribo) et aux questions de la sécurité / criminalité (Port-d'Espagne). Nous comparons ces connaissances avec la connaissance qualitative de ces lieux. Nous soutenons que Big Data est une ressource puissante pour découvrir des schémas intéressants, mais il faut être critique des mises en garde méthodologiques et tenir compte des spécificités socioculturelles de l'utilisation des technologies de l'information et de la communication.

European Journal of Development Research (2015) **27,** 505–522. doi:10.1057/ejdr.2015.48

Keywords: big data; Caribbean; inclusive development; mapping; social media; urban

Introduction

We live in an urbanizing world with more than half of the world population living in cities, as of 2008 (Martine and Marshall, 2007). Current and future urban growth is mainly situated in cities in the Global South and to a large extent in small (<500 000) and medium-sized (500 000–1 million) cities (UN-Habitat, 2013). These are simultaneously seen as nodes of opportunities, change and growth as well as concentrations of threats, inequality and vulnerability. Urban areas are locations of jobs, markets, education and health facilities, and of social change and cultural innovation. However, access to these is limited and unequal, creating concentrations of people living in extreme poverty. Moreover, impacts of climate change, crime and violence, political instability and (financial-) economic crises are threatening the lives of large populations living in urban environments and will further exacerbate urban inequalities and urban vulnerabilities.

Addressing such urban issues requires, among other things, adequate and relevant policies. Such policies develop as a result of power-loaded interactions between state, private sector and civil society actors operating at various scales (Torfing *et al*, 2012). In such governance processes, knowledge, data and information play a crucial role (Hordijk and Baud, 2006; Baud *et al*, 2009). One of the challenges for urban governance, in particular in the Global South, is the access and availability of relevant and reliable population data, for instance because of erroneous population projections (Cohen, 2006). Urban planners and decision makers require reliable and up-to-date information regarding, for example, population distributions and characteristics, environmental aspects, vulnerable groups and vulnerable locations. Conventional information sources relating to the urban population are the census covering the entire population or well-thought surveys carried out for a representative sample of the population. Environmental variables are often quantified by a combination of survey techniques and measurements, remote sensing data analysis or environmental modelling techniques, which are utilized less in cities in the Global South because of a lack of data, expertise or willingness (Pfeffer *et al*, 2013).

Recently, much attention has been given to Big Data as a new source of information and data. Big Data refer to large data sets of digitally born data, collected through human and non-human sensors. The data and information provided through these channels is commercially valuable, (for example, can be used to generate personalised advertisements), but its value in understanding social life and addressing societal issues is a topic of much recent research (for example, Kitchin, 2014). Big Data provide opportunities but also hold restrictions for policy development and implementation (Arribas-Bel, 2014; Taylor and Schröder, 2014). Applications in cities in the Global South are more related to disaster management and urban mobility, while applications in the Northern Hemisphere concern everyday management problems like traffic monitoring, incident management or measuring activities in transition and destination spaces (Sevtzu and Ratti, 2010; Steenbruggen *et al*, 2013a). In terms of poverty, Big Data applications mainly concern the global or national level (see for example, Noor *et al*, 2008; Smith-Clarke *et al*, 2014), with a preference for monetary approaches overlooking the multi-dimensionality of poverty.

The questions we pose here are: What value do Big Data hold for addressing urban questions beyond disasters and mobility in developing countries? What are the promises and constraints using such data? What do they add to conventional data, or can they replace them? Piotrowski (2014) considers Big Data sources rather as a complement to traditional data collection methods and analysis, and not so much as substituting these. In this paper, we explore the possibilities of Big Data through an analysis of tweets, focusing on two pressing issues: flooding and security/crime.

Hereafter, we first engage in the academic debate concerning Big Data. We address the variety of meanings of Big Data, and try to get an idea of the potential of Big Data for analysing, understanding and addressing pressing urban issues in less developed cities like those in the Caribbean, but also found elsewhere. Subsequently, we address the challenges and demands of Big Data for urban research and professional applications in that particular context. After this literature review we shift to our case studies. Specifically, we ask what an analysis of Twitter data in two Caribbean cities contributes to a better understanding of and adequate policy response to water-related risks (in Paramaribo, Suriname) and security/crime issues (in Port of Spain, Trinidad and Tobago). We conclude with forward thoughts on how to take advantage of the data revolution, while being reflective on its shortcoming for analysing pressing urban issues and its potential for inclusive development. We prefer data revolution to Big Data revolution because the former also recognises the value of traditional data sources such as structured surveys or qualitative data, by some also referred to as small data.

Background

Big Data: What is it and What Does it Mean?

While the notion of and term 'Big Data' has existed for a while (see for example, Press, 2013), it received a considerable push with the exponential growth of digital-born and networked data from a diversity of human and non-human sensors. On the one hand, there seems to be a common understanding of what Big Data means when we use the term. On the other, definitions vary across disciplines, contexts and authors. A non-exhaustive review of academic contributions concerning Big Data in geography, planning and related fields identifies a number of common characteristics building on the four classical V-properties of Big Data – volume, velocity, variety and veracity – the first three coined by Laney (2001). Specifically, Big Data are massive quantities of information (terabytes/petabytes) produced by and about people, things and their interactions (Boyd and Crawford, 2012; Kitchin, 2014), which are too large to fit into an Excel spread sheet (Batty, 2013) as it can only hold 65 536 rows and 256 columns. They are produced in real time or near real time (Kitchin, 2014), resulting in a fine temporal granularity. Furthermore, there is not one single type of data generator, but a variety of structured and unstructured active and passive means through which data are produced, putting stress on the reliability and accuracy of produced data because of the absence of well-thought sample schemes for guaranteeing representativeness and therefore a sound statistical analysis (Boyd and Crawford, 2012; Kitchin, 2013; Arribas-Bel, 2014). While these four Vs mainly deal with the data production side, additional characteristics concern the analytical part, such as the capacity to query, aggregate and link data sets (Boyd and Crawford, 2012, p. 663), with an emphasis on the 'linking' to draw relations between different data based on common fields, also referred to as 'vinculation' (Sui, AAG panel, 2014). Big data are also considered to be flexible and scalable, as they can be enriched with additional fields and expand in size rapidly. On the basis of the extent to which data production is steered, Kitchin (2013) classifies Big Data into *directed, automated* and *volunteer* data production, to distinguish whether data collection about an entity or place is operated or structured by a human being, often used for surveillance and monitoring purposes; whether it happens automatically because of automated functions, devices or systems leaving digital traces; or whether data is produced on a voluntary basis by engaging in social media or open source platforms, where data come into being by means of crowdsourcing activities. Typical examples often mentioned in relation to Big Data are, in the order of Kitchin's (2013) classification: health or government records, GPS data and sensor data; call logs of mobile phones, or other digital traces left by individuals through online behaviour, financial transactions or other digital activities; and data produced by activities on social media platforms such as Facebook, Twitter, Foursquare, Instagram or Flickr.

Promises, Potential and Challenges of Big Data Sources

Owing to their size, diversity, wide coverage and openness, but also because of their fine temporal and spatial granularity, Big Data are considered to be well suited for the study of complex dynamic systems such as cities, and to the analysis of their socio-spatial structure and character (Sevtsuk and Ratti, 2010; Cranshaw *et al*, 2012; Batty, 2013; Graham and Shelton, 2013; Arribas-Bel, 2014; Kitchin, 2014). For instance, attempts are made to classify urban areas by applying spectral clustering from the science toolbox to social data produced by means of social media platforms like Foursquare (foursquare.com), recording check-ins in real-time. This offers a glimpse on 'livehoods' by means of the 'consumption' of urban amenities in terms of where, what, how often and by whom (see also Cranshaw *et al*, 2012), assuming that people who are similar visit similar places. While the profiling of social groups or areas has been addressed in earlier studies on geodemographics (Harris *et al*, 2005), or clustering approaches using Census data (Martinez, 2005), the clustering of social media data – though only valid for a select group – may uncover unexpected patterns as the input is not pre-defined. Because of its near real-time data production, it has the capacity to map socio-spatial dynamics and human behaviour in city-scapes. The clustering of social media data also enriches our understanding of human interactions among people (social networks) and non-linear relations with the physical environment (for example, Lazer *et al*, 2014). Overall there are multiple research applications of social media data, and in particular Twitter (see for example, Bruns and Burgess, 2012; Bruns and Stieglitz, 2012; Moe, 2012; Takhteyev *et al*, 2012; Lewis *et al*, 2013), largely because of easy access.

However, there are also numerous applications of utilising mobile phone data to determine where people are (or have been) in case of humanitarian responses, to approximate human flows and crowding of people, or to predict human behaviour. Having knowledge about flows and the likely formation of human crowds assists aid agencies and transportation planners in designing preventive measures for large events, and in managing congestions of transportation systems. Measuring human flows by means of mobile phone traces is experimented with in transportation (Steenbruggen *et al*, 2013a) and migration studies (Taylor, submitted), but also in the spread of diseases related to migration patterns (Tatem *et al*, 2014). Recent cases of mobile phone applications include prediction of Malaria spread in Kenya and strategic malaria elimination planning in Namibia (Wesolowski *et al*, 2012; Tatem *et al*, 2014); humanitarian response after the Haiti earthquake (Heinzelmann and Waters, 2010); urban mobility in Rome (Sevtsuk and Ratti, 2010); transportation planning in Côte D'Ivoire (as part of the D4D initiative) or in the Netherlands (Steenbruggen *et al*, 2013b; Nanni *et al*, 2014); and understanding financial transactions and flows and price developments (UN Global Pulse, 2012), the latter mainly at national and global scale.

Big Data analytics are also adopted in the private and public sphere, with uses such as individual profiling and targeting, and the subsequent prediction of human behaviour through linking a variety of data produced by social media platforms, online activities, mobile phones or smart cards. Well-known applications are marketing strategies such as sending advertisements based on people's search and activity profile; assisting companies in the development of their locational strategy and range of products offered; forecasting political elections; and the performance monitoring of individuals and organisations. Emergent and rapidly developing practices can also be found in the domains of governance, politics and surveillance (Townsend, 2013; Tufekci, 2014). As such, Big Data offer a possible avenue for engineering a better world (Eagle and Greene, 2014, p. 2).

Apart from these more applied applications, there are also claims that Big Data analytics offers new opportunities for research; in particular, to gain insights into new phenomena or in generating new questions, but also by revisiting former research questions in a new way (Arribas-Bel, 2014). It may also open new avenues for the spatial sciences, given that many data sources have spatial attributes or accurate geographical coordinates (Kitchin, 2013).

With respect to data as such, recent discussions concern the potential of Big Data to substitute traditional data sources because of their wider geographical coverage and 'real-time' character (Arribas-Bel, 2014). Traditional data collection, such as the regular Census, is carried out every 10 years, and therefore loses its validity over time (especially in fast-growing and dynamic cites). Survey data are collected more frequently, but from a smaller demographic or geographical sample. Moreover, many cities in the Global South still lack accurate and up-to-date databases and the capacity to utilise available data for governing their resources and entities in a strategic manner. Recent debates suggest that Big Data could fill these data gaps (Kitchin, 2013; Arribas-Bel, 2014). However, the typical Big Data sources as outlined above are not purposely produced for exploring a particular urban issue; they are, rather, a 'side effect' (Arribas-Bel, 2014), and therefore lack quality in terms of how an issue is framed and then measured, and whether the data set is representative for that particular issue.

Despite the promising potential, Big Data and Big Data analytics come with challenges. A first challenge in utilising Big Data – here focused on social media data or call data records (CDR) from mobile phones – for (urban) research concerns their *representational validity* in terms of the demographic and social-economic profile and the geographic spread of the contributors on an automated or voluntary basis, as well as their locational accuracy (for example, Boyd and Crawford, 2012; Haklay, 2013; Lazer *et al*, 2014). First, researchers may not have access to the complete database, but only to an arbitrary sample independent from the research problem they are examining. Twitter, for instance, offers different sample sizes, ranging from 1 to 10 per cent (Bruns and Liang, 2012). People also make choices in terms of social media platforms, mobile phone providers and the use of smart cards, and therefore several data-producing applications such as Foursquare are biased towards smartphone users or users of other mobile wireless devices. Second, digital-born data only relate to those active on social media platforms, or – in the case of mobile phones – those using a mobile phone actively. Thus, offline people or groups (for reasons including digital illiteracy, lack of infrastructure, lack of financial means, language, unfamiliarity, refusal to participate in the social media and mobile hype), or offline activities, are missing from this rich data landscape (Crutcher and Zook, 2009; Graham, 2011; Cranshaw *et al*, 2012). Moreover, mobile phone data from the same phone can refer to multiple people, or an individual can own multiple phones, making interpretation difficult (Taylor, submitted). Third, if not GPS-enabled, location is assigned in a post-processing procedure (illustrated in Steenbruggen *et al*, 2013b) for the 'geocoding' of CDR, subject to an unknown uncertainty in geographic location. On a highway, a signal may belong to a car driver, while in an urban setting multiple producers are possible (resident, pedestrian, cyclist or motorized driver). Fourth, CDR, to the extent accessible to the research community, are only offered in aggregate form, as privacy regulations (if in place, as is the case in many Western countries) may prohibit the sharing of individual data records. Finally, due to the semi-open access nature of automated or volunteer data generation, it is obvious that these are not generated in a well-thought controlled experiment guaranteeing replicability, and therefore not suited to test hypotheses as done in conventional statistical analysis, but subject to exploratory research designs. Accordingly, conclusions can only be drawn on the basis of the collective behaviour studied while being aware of the (geographic) shortcomings. To account for the representational validity, research findings need to be triangulated with other data sources to examine the reasons for the voids on a map (Graham and Shelton, 2013). A promising example, though not necessarily applicable to the urban environment, is the integration of mobile phone data, surveillance data and satellite imagery to explore the spread of diseases (see for example, Tatem *et al*, 2014).

The second main challenge in analysing Big Data sources is of a technical nature. In particular, if large volumes of data need to be analysed or if the aim is to link different data sources to find emerging patterns, social scientists or urban practitioners often lack the necessary technical and coding skills that data scientists are taught (Kitchin, 2013). These skills are required to extract data from the variety of data-rich platforms like the Internet or social media applications. While APIs (application programming interfaces) are designed to support data extraction, they still require some technical know-how in how to read, change and run codes (which are often written in the widely used interpretive Python language), and to access and extract the desired information (for example, Arribas-Bel, 2014). It should be noted that access may not only be restricted due to the lack of technical skills, but also because data producers like mobile phone companies consider their data a competitive advantage, limiting their willingness to share them with ordinary users, thereby challenging existing power relations. Furthermore, handling and analysing these massive unstructured data sets requires advanced computational methods, skills and computing power, which do also not belong to the standard equipment of social scientists. Examples of the skills required are machine learning to train the computer to detect patterns; knowledge discovery methods; spectral clustering methods or visualisation techniques (see for example, Cranshaw *et al*, 2012; Arribas-Bel, 2014) – the latter, for instance, implemented in Gapminder (gapminder.org), Flowminder (flowminder.org) and Digital OnLine Life and You (DOLLY; Floating.Sheep, n.d.) – or the monitoring tools developed by Global Pulse (UN Global Pulse, 2014). UN Global Pulse applied content analysis to Twitter data at the national level. One needs to be aware that words chosen for searching tweets are not attuned to the local context, and that the knowledge on penetration of a particular social media platform is not incorporated in this analysis. It is not sensitive to scale, therefore hiding sub-national differences, and does not distinguish urban and rural contexts.

Many of the more geographic Big Data studies are carried out at the national or global scale. Urban studies mainly deal with understanding the human behaviour of city dwellers and the socio-spatial structure they produce, as well as transportation planning applications (for example, Batty, 2013). Of note here is that the potential of Big Data is not only seen within the research community, but – perhaps especially – outside or on its borders, with Big Data analytics also appearing in the public sphere. The Smart City movement is gaining traction in urban governance around the globe, attempting to apply insights gained from the analysis of Big Data to urban environments – engineering a better, smarter city in the process, and offering the necessary solutions for sustainable cities (cf. Kitchin, 2011; Batty *et al*, 2012; IBM, 2012; Townsend, 2013; Kitchin, 2014; Shelton *et al*, 2015 for a critique). Big Data is also emerging in the political agenda of cities in the Global South, for instance the planning of 100 Smart Cities in India (Rai, 2014).

Specific barriers in the use of Big Data in urban governance (and in particular for cities in the Global South) are data privacy issues, data access, human capacity and IT infrastructure, because the handling and use of Big Data requires statistical and technical skills and expertise. Civil servants may already find it challenging to query databases with GIS-based information on their city, and to interpret the resulting plots, and even more so the raw data handling and processing (Pfeffer *et al*, 2012). In order to deal with the Big Data challenges, governments need to collaborate with private sector firms, with the risk of ending up dependent upon this relationship (Baud *et al*, 2013). One way of moving forward would be to synchronise and link existing government databases and combine them with new data sources; however, interoperability is still a challenging issue even with traditional data sources, even more so in cities with less developed digital databases (Richter, 2014). Thus, while there would be the technological means to augment existing traditional data sources with (near) real time information, local conditions and capacities affect the realisation of the Big Data potential in practice, often creating new divides (Batty *et al*, 2012).

Methodology

In order to examine the potential of Big Data as a methodological tool for developing new knowledge on urban issues that matter in certain cities, eventually assisting those working towards inclusive development, we have selected two Caribbean cities: Paramaribo, Suriname, and Port of Spain, Trinidad and Tobago. The choice was based on the following criteria: size, focusing on smaller cities; exposure to a pressing issue, like security, an increased risk of flooding because of climate change, inequality, or poverty (see for instance, McGranahan *et al*, 2007); and familiarity with the local context. For each city we selected one pressing issue; flooding-related issues in Paramaribo, the city being at risk of heavier and shorter rainfalls in the near future because of climate change (Baud *et al*, under review); and security in Port of Spain, suffering severe violence.

Paramaribo is Suriname's main political, economic, social, administrative and residential centre, home to 240 914 residents, nearly 45 per cent of the total population of Suriname (541 638) (ABS, 2013). It is a sprawling city, situated about 20 km from the mouth of the Suriname River, bordering the Atlantic Ocean. Residential plots are relatively large and multi-storey buildings are scarce. Planning is largely absent, and Paramaribo develops in an ad-hoc, unstructured manner. About 90 per cent of Suriname's population, including Paramaribo, is located in a low-elevation coastal zone (McGranahan *et al*, 2007). The city regularly floods because of an inadequate run-off system and (combinations of) extreme rainfall and high tide. Climate change and over-reliance on the existing infrastructure are expected to further increase flood risks. ICT and mobile phone penetration data are only available at country level. In 2013, 179 250 Surinamese (32 per cent) used Internet (aligned with the Caribbean average) and 99 820 had a Facebook account (Internet World Stats (2013)). Mobile phone penetration in 2013 was 127 per cent (UN Data (2014)).

Port of Spain, Trinidad and Tobago's capital city, is part of the heavily urbanised East-West corridor that stretches from Chaguanas (in the west of the country) to Arima (in the centre of the country) and has approximately 600 000 residents. While the administrative areas of Port of Spain account for 37 000 residents only, this number rises to 300 000 if seen as a city-region together with its adjacent city corporations Diego Martin and San Juan/Laventille (CSO, 2012). Over the last decade crime rates in Trinidad and Tobago have increased rapidly: from 118 murders in 2000, to 550 in 2008, to 405 in 2013 (www.ttcrime.com). Half of the murders are because of gang warfare and take place in the six neighbourhoods surrounding Port of Spain. In addition, violent crimes such as rape, robbery and burglary are omnipresent (TT Crime (2014)). The murders, a few years of high kidnapping rates and rising firearms crimes create a low sense of security (26 per cent) in the country, the lowest among seven Caribbean cases (Zimmermann *et al*, 2012, p. 19). ICT and mobile phone penetration data are only available at country level. In 2013, 656 611 (53.1 per cent) Trinidad and Tobagonians used the Internet (more than the Caribbean average) and 484 780 had a Facebook account (Internet World Stats (2013)). In 2013, the mobile phone penetration rate was 145 per cent (UN Data (2014)).

For both cities, we analysed all geocoded Twitter data collected for the year 2012, made available by the project team of DOLLY (Floating.Sheep, n.d.; Crampton *et al*, 2013), affiliated with the University of Kentucky. Although real-time Twitter data can be accessed easily via its API (dev.twitter.com/) or acquired commercially via third-party vendors such as Gnip, DOLLY provides academic researchers with a repository of billions of geo-located Tweets dating back to June 2012. It allows for real-time research and analysis, and is therefore considered well-suited for an exploration of the capacity of social media data, specifically derived from Twitter, for urban analysis. Nevertheless, to date it is not yet entirely open access because of the dependence on legal agreements with Twitter.

The geocoded Twitter data extracted from DOLLY covered the following bounding boxes: Paramaribo – 5.675434, −55.386887; 5.968485, −54.967346; and Port of Spain – 10.512455, −61.727028; 10.801944, −61.307487. All tweets within the city boundaries (as defined by the administrative geography boundaries for which GIS maps were acquired through our local partners) were selected for further analysis. This resulted in 180 941 tweets for Paramaribo and 109 940 for Port of Spain. In order to take into account the border effect, tweets 1 km beyond the administrative boundary were also selected for further inspection. Both for flooding and security, we developed a list of search terms that relate to water-related risks or security, in the given local context. For water-related risks we explored the natural factors contributing to this risk, like flooding. With respect to security, we investigated both perceptions of security and actual experience of crime, for example places that are named as being unsafe, and the mention of security issues such as robbery, theft and murder. (see Tables 3 and 6 for the terms that we could trace). The choice of search terms was also determined by the authors' local knowledge acquired in former research projects in these areas (for example, Baud *et al*, under review; Verrest, 2013). The resulting codes served to select those tweets that could potentially say something meaningful in relation to the issue explored. The result of the automated query was further inspected manually to remove irrelevant tweets, reducing them to 880 tweets for Paramaribo and 706 for Port of Spain. The statistical package SPSS 20 and the geographical information system (GIS) ArcGIS 10.1 were used to compute frequencies of the subsets by code and administrative area, to create case summaries for further qualitative examination, and to spatially map both the overall tweet density and the issue-related density, by aggregating the number of tweets to grid cells of 100 by 100 m.

The analysis was carried out at two levels – the city as a whole as delineated by its administrative boundaries, and at neighbourhood level (as utilised by the census). A more detailed interactive analysis making use of the Openstreetmap (openstreetmap.org) was done in those areas that showed high numbers of tweets with respect to the identified issues.

Analysis and Results

Paramaribo

There is Twitter activity throughout the city as shown in Figure 1, with high activities in the better-off areas or around public spaces and business and industrial areas. The clustering of higher densities in Flora, and in the western part of Beekhuizen, refers to specific institutions like schools or the city airport. Poor areas such as Pontbuiten and the less populated resort Livorno display low tweet activities.

The low number of water-related tweets at neighbourhood level (from the 180 941 tweets produced during 2012, only 880 have some mention of water-related issues see table 3) follows the general pattern: many water-related tweets mean high Twitter activity in general (see Tables 1 and 2). Most of these tweets are produced in Blauwgrond, Centrum and Rainville. These areas are

Figure 1: Tweet density in Paramaribo 2012; source data extracted from DOLLY, May 15 2014.
Sources: DOLLY, 2014; Census Office, 2004. Conception and design: Karin Pfeffer.

prone to water-related risks; however, residents here generally have means to adapt to this (Linnekamp *et al*, 2012). In our exploratory analysis we were also expecting to find tweets on water-related risks from resorts like Pontbuiten and Latour, both vulnerable to flooding, but whose residents belong to low-income groups and therefore do not have the means to cope with that risk. There is little reference to water-related issues as overall tweet density in these two neighbourhoods is low, 1.7 and 4.9 per cent, respectively (Tables 1 to 2).

Analysing the content of the tweets reveals that the large majority says something about rain, and only a few mention that streets are prone to flooding (commonly known as a major issue). There is no systematic pattern of streets referred to as flooded, although some specifically mention Keizerstraat in the city centre. A qualitative analysis of the tweets reveals emotional relations with weather situations, illustrated with few examples: ‘Rain Rain Rain. This weather just makes you lazy.’ Or ‘Home … brrrr cold rainy day’.

Port of Spain

Just as in Paramaribo, in Port of Spain tweets are produced throughout the city, with higher occurrences in commercial areas (for example, Woodbrook and Port of Spain proper) and a few in low-income areas such as East Port of Spain, Belmont and Gonzales (Figure 2, Table 4).

From the 109 940 tweets, only 706 contain security-related issues, of which the majority feature the words police, safe, murder, crime/criminal and shot/shooting. In most areas this is below the average of 0.64 per cent of the tweets Tables 5 to 6. In East Port of Spain, with a

Table 1: Distribution of tweets across resorts, population characteristics and type of area

Resort	*Frequency*	*%*	*Cum. %*	*People*	*Type of area*
Beekhuizen	12 022	6.6	6.6	17 185	Lower middle-class mixed residential/industrial
Blauwgrond	34 693	19.2	25.8	31 483	Residential well off
Centrum	26 798	14.8	40.6	20 631	Commercial / low income
Flora	21 362	11.8	52.4	19 538	Residential / mixed income
Latour	8809	4.9	57.3	29 526	Residential / low income
Livorno	2733	1.5	58.8	8209	Residential/ industrial/low income
Munder	8998	5.0	63.8	17 234	Residential / middle income
Pontbuiten	3003	1.7	65.4	23 211	Residential / low income
Rainville	19 802	10.9	76.4	22 747	Residential / middle income
Tammenga	15 329	8.5	84.9	15 819	Residential/ middle income
Weg naar Zee	6767	3.7	88.6	16 037	Semi-residential / middle income
Welgelegen	20 625	11.4	100.0	19 304	Residential / high income
Total	180 941	100.0	—	—	—

Source: DOLLY, 2014: geocoded Twitter sample for 2012.

Table 2: Frequencies of security- related issues by resort

Resort/category	*Water*	*Rain*	*Storm*	*Flash*	*Total*
Beekhuizen	2	80	0	1	83
Blauwgrond	7	**172**	2	5	186
Centrum	8	**108**	7	3	126
Flora	6	79	3	2	90
Latour	0	51	3	0	54
Livorno	0	12	0	0	12
Munder	0	32	5	0	37
Pontbuiten	1	11	0	0	12
Rainville	5	**101**	2	0	108
Tammenga	1	66	0	0	67
Weg naar Zee	1	24	1	1	27
Welgelegen	6	66	4	2	78
Total	37	802	27	14	880

Source: DOLLY, 2014: geocoded Twitter sample for 2012.

relatively low overall tweet activity (2.6 per cent), nearly 1 per cent are related to safe and police, while in St Clair, contributing only 7.5 per cent to the overall tweet activity, many tweets (1.75 per cent) are related to security, specifically crime, police and murder. However, these are mainly produced by one individual, a radio DJ spreading newsflashes. As expected, security-related tweets cluster in the centre where social and commercial facilities and activities concentrate. Most security-related tweets are produced in the central business district, shopping areas, hotels and crowded public spaces, areas where people work, study, shop, go out, dance, party and drink and that together produce more than 50 per cent of all tweets (that is, Port of Spain proper and Woodbrook). The tweet concentration beyond the northern city boundaries shows that it is not limited to the main city only, rather that it is a cross-boundary phenomenon. Figure 3 illustrates these spatial patterns and Figure 4 displays the percentage of security-related tweets with respect to all tweets. The map suggests that there are a few hotspots where the number of

Table 3: Frequencies by code, based on tweets within city boundaries

Category	*Frequency*	*%*	*Cum. %*
Rain	802	91.2	91.1
Water	37	4.2	95.3
Storm	27	3.1	98.4
Flash	14	1.6	100.0
Total	880	100	—

Source: DOLLY, 2014: geocoded Twitter sample for 2012.

Table 4: Distribution of tweets, number of people (2000), dominant land use type by area

Census. area	*Frequency*	*%*	*Cum. %*	*People*	*Type*
Belmont	8821	8.0	8.0	**11 627**	Residential / mixed income
Cocorite	1582	1.4	9.5	1907	Residential / low income
East Port of Spain	3026	2.8	12.2	11 681	Residential / low income
Ellerslie Park	1535	1.4	13.6	386	Residential / middle income
Federation Park	561	0.5	14.1	328	Residential / middle income
Gonzales	1498	1.4	15.5	2811	Residential / low income
Long Circular	411	0.4	15.9	435	Residential / low income
Newtown	4461	4.1	19.9	1114	Commercial / mixed income
Port of Spain Port Area	3260	3.0	22.9	9	Industrial
Port of Spain Proper	**36 915**	33.6	56.5	4316	Commercial / low income
Sealots	339	0.3	56.8	1859	Residential / very low income
St Clair	8200	7.5	64.2	595	Residential / high income
St James	14 557	13.2	77.5	6966	Commercial / mixed income
Woodbrook	**24 774**	22.5	100.0	4480	Commercial / middle income
Total	109 940	100.0	—	—	—

Source: DOLLY, 2014: geocoded Twitter sample for 2012.

security-related tweets is 50 or close to 50 per cent. However, a closer examination of the darker locations reveals low tweet activity in general, with not more than 8 tweets for the whole year 2012.

Discussion and Conclusion

Big Data generate much attention in academic and policy circles, and are considered valuable means in the detection of societal problems and the development of relevant policy solutions (for example, Delgado, 2014). Our interest is in the relevance of Big Data in addressing pressing urban issues in the Global South. The governing actors of these cities are confronted with a multitude of vulnerabilities and threats, such as fast and informal expansion of the city, natural disasters, social inequality, political unrest and violence, and a lack of accurate data on these issues. An important question to examine is therefore how Big Data may provide useful information to develop adequate responses to some of these issues. In this article we have explored how one type of Big Data, that is, tweets sent through Twitter, can be useful in addressing two such pressing issues: floods and security. We have explored this for two small (<500 000 inhabitants) capital cities in the Caribbean: Paramaribo, Suriname, susceptible to

Table 5: Frequencies of security-related issues by census area

Category	*Bel*	*Coc*	*East POS*	*Ell. Park*	*Newt.*	*POS port*	*POS proper*	*Seal.*	*St. Cl.*	*St. Jam.*	*Woodbr.*	*Total*
Security	0	0	0	0	0	0	1	0	12	0	0	13
Crime	1	1	1	0	3	2	11	0	**25**	11	9	64
criminal	0	1	0	0	1	0	4	0	6	4	7	23
Violence	0	0	0	0	0	0	0	0	0	0	2	2
Fear	1	1	0	0	0	0	0	0	1	0	0	3
Gun	0	0	0	0	0	0	22	0	6	0	22	50
Weapon	0	0	0	0	0	0	1	0	0	0	1	2
fight	1	0	0	0	0	1	9	0	2	1	1	15
Police	5	4	9	2	0	2	**47**	0	**35**	**32**	**60**	196
safe	11	0	12	0	5	1	**64**	1	12	11	18	135
Danger	1	0	0	0	0	0	0	0	0	0	1	2
Murder	0	1	1	0	1	2	19	0	**25**	2	**28**	79
shot	2	3	4	2	1	0	13	0	12	1	19	57
shooting	0	1	2	0	0	0	2	0	2	0	8	15
Gang	2	0	0	2	0	0	3	0	6	3	**27**	43
Bad boy	0	1	0	0	1	0	0	0	0	0	1	3
aggressive	1	1	0	0	0	0	0	0	1	0	1	4
	25	14	29	6	12	8	**196**	1	**145**	65	**205**	706

Source: DOLLY, 2014: geocoded Twitter sample for 2012.

Table 6: Frequencies by code, based on tweets within city boundaries

Category	*Frequency*	*%*	*Cum. %*
Police	196	27.8	27.8
safe	135	19.1	46.9
Murder	79	11.2	58.1
Crime	64	9.1	67.1
shot	57	8.1	75.2
Gun	50	7.1	82.3
Gang	43	6.1	88.4
Criminal	23	3.3	91.6
fight	15	2.1	93.8
shooting	15	2.1	95.9
Security	13	1.8	97.7
aggressive	4	0.6	98.3
Fear	3	0.4	98.7
Bad boy	3	0.4	99.2
Violence	2	0.3	99.4
Weapon	2	0.3	99.7
Danger	2	0.3	100.0
Total	706	100.0	—

Source: DOLLY, 2014: geocoded Twitter sample for 2012.

regular flooding, and Port of Spain, Trinidad and Tobago, experiencing high rates of crime and feelings of insecurity. We conducted a spatial analysis of the number and distribution of tweets in both cities, the number and distribution of tweets related to water or security, the type of information that is spread through sending a tweet and, finally, explanations for spatial concentrations of water- or crime-related tweets.

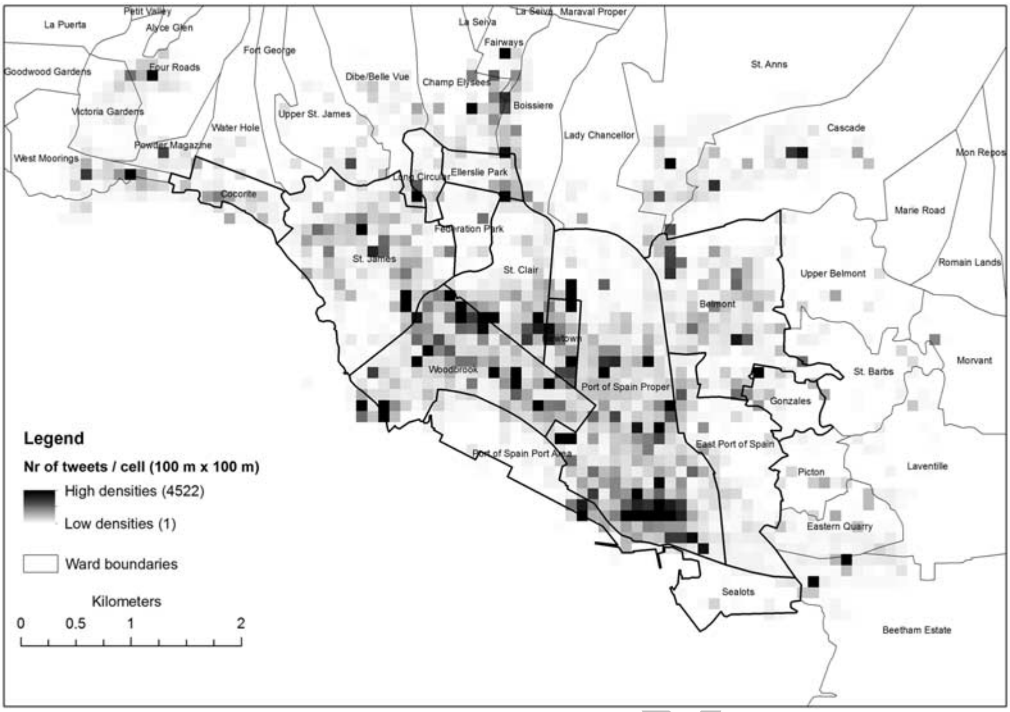

Figure 2: Tweet density in Port of Spain 2012; source data extracted from DOLLY, May 15 2014. *Sources*: DOLLY, 2014; Census Trinidad, 2000; Conception and design: Karin Pfeffer.

For our analysis we had access to 109 940 tweets in Port of Spain and 180 941 in Paramaribo. In both cities, the sending of tweets is concentrated in business and commercial districts, around institutions like educational facilities or transportation infrastructure. Furthermore, fewer tweets are sent from low-income areas such as Pontbuiten and Latour in Paramaribo, and East Port of Spain and Laventille in Port of Spain. Of note here is that the location from where a tweet is sent is not necessarily the home location; tweet patterns such as densities pertain to geographic areas, and not to people living in these areas. A filtering of only those tweets that relate to water in Paramaribo (n=880) and security in Port of Spain (n=706) makes clear that less than 1 per cent of all tweets are related to the pressing issues considered here. The distribution of the issue-related tweets follows the general distribution of tweets. Possible explanations for these variations in distributions are Internet access and quality, with commercial and business districts better equipped than residential areas, and low-income groups experiencing poorer access than high-income groups. Our findings confirm general understandings in the literature, namely, that examining tweets (and other Big Data) is subject to several methodological biases and leaves specific social groups and geographical areas underexposed. The social-spatial fragmentation of many cities in the Global South further increases these biases. Hence, developing policies based on tweets may actually further increase inequality as specific groups and areas may be left underexposed in policies and Smart City applications (for example, Townsend, 2013).

Floods and crime are pressing issues, but neither exclusively involves low-income groups or areas. In Port of Spain, gang-related violence is mostly concentrated in low-income areas surrounding the administrative boundaries of the city, but other forms of crime and perceptions of (un)safety are not concentrated in specific areas. Similarly, floods in Paramaribo are common

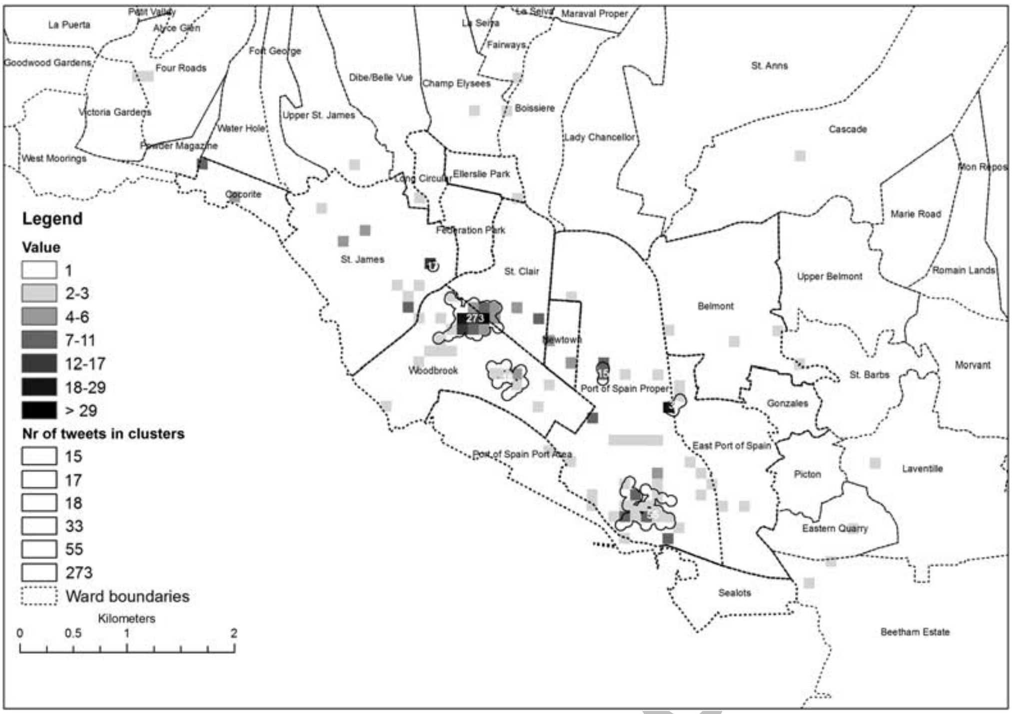

Figure 3: Tweet density in Port of Spain of security-related issues 2012; source data extracted from DOLLY, May 15 2014.
Sources: DOLLY, 2014; Census Trinidad, 2000; Conception and design: Karin Pfeffer.

both in Paramaribo North, an area with many high-income neighbourhoods, and Paramaribo South, where low-income neighbourhoods are concentrated. Hence, we would expect that despite biases in people who send tweets and locations from where tweets are sent, potential valuable information regarding floods and crime may be tweeted. Our content analysis of the tweets shows that most tweets express a mood ('It is raining: I feel like staying in bed') or opinion ('poverty is the parent of revolution and crime') or general expression ('stay safe'). One of the potentials of social media data is the production of very fine temporal and spatial granularity. Such data may provide detailed information on areas that are experienced as unsafe, or that are susceptible to floods, not necessarily the home locations. Moreover, such data can stimulate an immediate response, for example by the police. In the cities we consider, however, very few of the tweets actually express real-time experiences with floods or crime/violence. As such, the value of tweets for the development of immediate responses or long-term policies to manage violence in Port of Spain and floods in Paramaribo is very limited. Tweets do not bring up information regarding real-time flood experiences, or real-time perceptions of safety or violence.

These findings suggest that Big Data such as Twitter may not provide relevant data for policy development in Port of Spain and Paramaribo. Twitter's limited value is explained by the question 'Who is tweeting and from where?', but also by the content of tweets and the tweeting networks. There are no globally applicable answers to these questions, due to the situated local context. Assessing the characteristics of local users and local use of specific social media is therefore a necessary step before developing instruments to use Big Data produced by social media for policy development. As such, the global approach taken by promising initiatives such as Global Pulse may be inadequate. In addition, while Big Data are

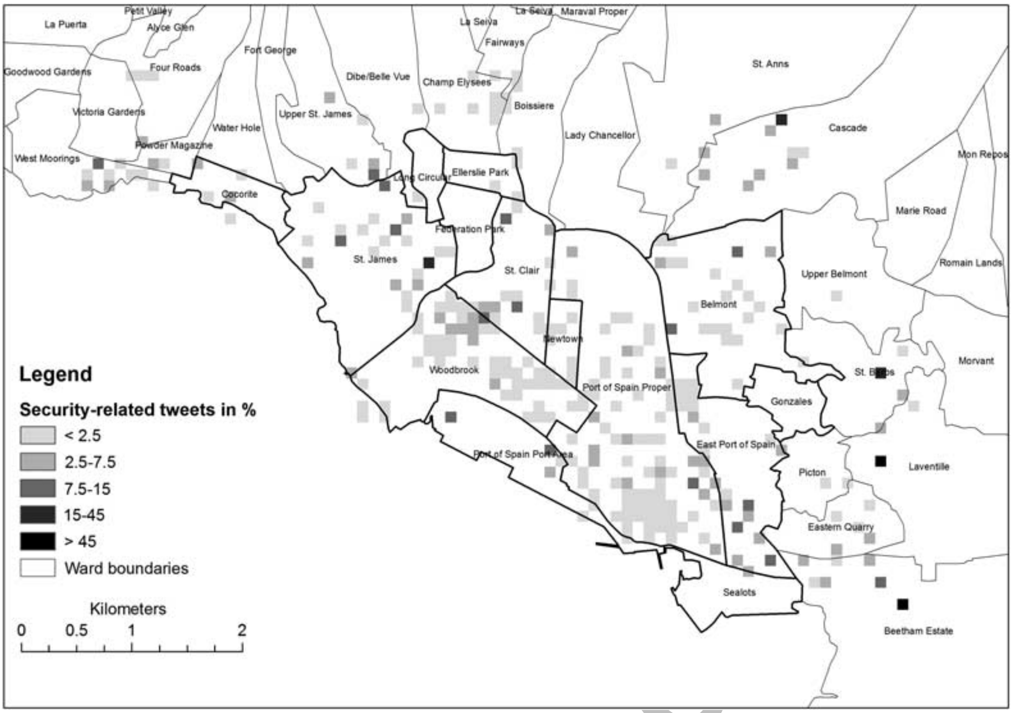

Figure 4: Percentage of security-related tweets in Port of Spain 2012; source data extracted from DOLLY, May 15 2014.
Sources: DOLLY, 2014; Census Trinidad, 2000; Conception and design: Karin Pfeffer.

interesting in themselves, combining them with more traditional data sources such as systematic surveys, qualitative sources or controlled measurements will deliver a more comprehensive picture for the given research problem. However, this requires strategic collaborations to account for the lack of skills, capacities and resources for data analytics in Southern contexts.

The discussion on Big Data is relevant for inclusive development. Big Data is expected to provide opportunities to expand knowledge on a wide variety of development indicators (covering several dimensions), and on potential crises and changes in behaviour resulting from this. Mobile phone data, for example, is used in disaster response, to estimate poverty, or to assess the spread of infectious diseases (Heinzelmann and Waters, 2010; Wesolowski *et al*, 2012; Smith-Clarke *et al*, 2014). Such data and information could assist governments, NGOs and policymakers in creating appropriate policies and developing effective interventions to target populations in strategic locations (UN Global Pulse, 2012). However, in order to do so, it is crucial to link the analysis to the local social-economic and cultural conditions and resources (ibid). Moreover, the collaboration between development practitioners and data scientists is essential to produce and use the information and analysis needed. In other words, '… it must be locally meaningful, must draw on local resources, and must demonstrate benefits to local residents as well as [city] authorities' (Taylor, 2014, p. 3). A number of challenges and concerns pertain to Big Data and inclusive development. The first is related to the required technical capacity to analyse Big Data (UN Global Pulse, 2012, p. 7). In terms of storage and technologies, the fast developments in this domain and the potential costs involved may further increase the digital divide and create new divides (Batty *et al*, 2012, p. 485). Two other concerns are characteristic of the data themselves. The first pertains to privacy and ethical issues of collecting this kind of data. Being a global problem, it is relevant especially in the Global South, as legal frameworks protecting residents are much less developed there (Taylor and Schröder, 2014). Second, Big Data tend to collect data of users of (specific forms of) digital devices and technologies, these being gender, age and income specific. If policy development is based on such data analysis, it may only reflect the concerns of subsets of the population and lead to ineffective policies and increasing inequality (ibid). While algorithmic manipulation may have positive outcomes for tracking diseases, controlling traffic flows, informing on market prices or making our everyday lives easier, it will also engineer negative outcomes such as rising food prices or exclusion of targeted populations. Moreover, surveillance, individual profiling and having the privilege of access to such databases make those in power more powerful, and may lead to unforeseen situations such as higher food prices, dropping real estate values, even higher levels of inequality, and shifting power asymmetries in urban governance networks (UN Global Pulse, 2012; Taylor *et al*, 2014; Tufekci, 2014).

References

ABS (2013) *8ste volks- en woningtelling in Suriname (vol 1): Demografische en sociale karakteristieken en migratie*. Paramaribo, Suriname: ABS.

Arribas-Bel, D. (2014) Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography* 49: 45–53.

Batty, M. (2013) Big data, smart cities and city planning. *Dialogues in Human Geography* 3(3): 274–279.

Batty, M. *et al.* (2012) Smart cities of the future. *The European Physical Journal Special Topics* 214(1): 481–518.

Baud, I., Verrest, H., Eleftheriadou, E., Muiderman, K. and Van der Staak, K. (under review) Building adaptive capacity to climate change and reducing urban injustices in the Southern Caribbean: Risk perceptions, policies, and implementation in fragmented governance arrangements. Geoforum, submitted.

Baud, I. *et al.* (2013) The development of Kalyan Dombivili: Fringe city in a metropolitan region. City report. Bonn: EADI – chance2sustain.

Baud, I., Pfeffer, K., Sridharan, N. and Nainan, N. (2009) Matching deprivation mapping to urban governance in three Indian mega-cities. *Habitat International* 33(4): 365–377.

Boyd, D. and Crawford, K. (2012) Critical questions for big data. *Information, Communication & Society* 15(5): 662–679.

Bruns, A. and Burgess, J. (2012) Researching news discussion on twitter. *Journalism Studies* 13(5–6): 801–814.

Bruns, A. and Liang, Y. (2012) Tools and methods for capturing twitter data during natural disasters. *First Monday* 17(4), 2 April.

Bruns, A. and Stieglitz, S. (2012) Quantitative approaches to comparing communication patterns on twitter. *Journal of Technology in Human Services* 30(3–4): 160–185.

Census Office (2004) Districtsresultaten 1: Paramaribo. Paramaribo, Algemeen Bureau voor de Statistiek. Paramaribo, Algemeen Bureau voor de Statistiek.

Cohen, B. (2006) Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. *Technology in Society* 28(1): 63–80.

Crampton, J.W. *et al.* (2013) Beyond the geotag: Situating big data and leveraging the potential of the geoweb. *Cartography and Geographic Information Science* 40(2): 130–139.

Cranshaw, J., Schwartz, R., Hong, J.I. and Sadeh, N.M. (2012) *The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City.*. Palo Alto, California: ICWSM, The AAAI Press.

Crutcher, M. and Zook, M. (2009) Placemarks and waterlines: Racialized cyberscapes in post-Katrina google earth. *Geoforum* 40(4): 523–534.

CSO (2012) Trinidad and Tobago 2011 Population and Housing Census: Demographic Report. Port of Spain: CSO.

Delgado, R. (2014) Lifting up: How big data can help eliminate poverty. Blog posted on 23 May, https://smartdatacollective.com/rick-delgado/200566/lifting-how-big-data-can-help-eliminate-poverty, accessed 11 June 2014.

Eagle, N. and Greene, K. (2014) *Reality Mining*. Cambridge, MA; London: MIT Press.
Floating.Sheep (n.d.) DOLLY, http://www.floatingsheep.org/p/dolly.html, accessed 24 October 2014.
Graham, M. (2011) Time machines and virtual portals: The spatialities of the digital divide. *Progress in Development Studies* 11(3): 211–227.
Graham, M. and Shelton, T. (2013) Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography* 3(3): 255–261.
Haklay, M. (2013) Neogeography and the delusion of democratisation. *Environment and Planning A* 45(1): 55–69.
Harris, R., Sleight, P. and Webber, R. (2005) *Geodemographics, GIS and Neighbourhood Targeting*. Chichester: Wiley.
Heinzelman, J. and Waters, C. (2010) *Crowdsourcing Crisis Information in Disaster-Affected Haiti*. Washington DC: U.S. Institute of Peace.
Hordijk, M. and Baud, I. (2006) The role of research and knowledge generation in collective action and urban governance: How can researchers act as catalysts? *Habitat International* 30(3): 668–689, doi:10.1016/j.habitatint.2005.04.002.
IBM (2012) How to transform a city: Lessons from the IBM smarter cities challenge. IBM Smarter Cities White Paper, http://asmarterplanet.com/files/2012/11/Smarter-Cities-WhitePaper_031412b.pdf, accessed 24 October 2014.
Internet World Stats (2013) Internet usage, facebook subscribers and population statistics for all the Americas world region countries 31 December, http://www.internetworldstats.com/stats2.htm#americas, accessed 20 June 2014.
Kitchin, R. (2011) The programmable city. *Environment and Planning B: Planning and Design* 38(6): 945–951.
Kitchin, R. (2013) Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267.
Kitchin, R. (2014) The real-time city? Big data and smart urbanism. *Geojournal* 79(1): 1–14, Knowledge management tools. *Information, Communication & Society*, 16(2): 258–285.
Laney, D. (2001) *3D Data Management: Controlling Data Volume, Velocity and Variety*. Stamford, CT: Meta Group.
Lazer, D., Kennedy, R., King, G. and Vespignani, A. (2014) The parable of google flu: Traps in big data analysis. *Science* 343(6176): 1203–1205.
Lewis, S.C., Zamith, R. and Hermida, A. (2013) Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media* 57(1): 34–52.
Linnekamp, F., Koedam, A. and Baud, I.S.A. (2013) Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media* 35(3): 447–456.
Martine, G. and Marshall, A. (2007) *State of World Population 2007: Unleashing the Potential of Urban Growth,* UNFPA.
Martinez, J.A. (2005) Monitoring intra-urban inequalities with GIS-based indicators – With a case study in Rosario, Argentina. PhD thesis. Utrecht, The Netherlands: Utrecht University/ITC.
McGranahan, G., Balk, D. and Anderson, B. (2007) The rising tide: Assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and Urbanization* 19(1): 17–37.
Moe, H. (2012) Who participates and how? Twitter as an arena for public debate about the data retention directive in Norway. *International Journal of Communication* 6: 1222–1244.
Nanni, M. *et al.* (2014) Transportation planning based on GSM traces: A case study on Ivory Coast. In: J. Nin and D. Villatoro (eds.), Springer International Publishing, Cham: CH, pp. 15–25.
Noor, A.M., Alegana, V.A., Gething, P.W., Tatem, A.J. and Snow, R.W. (2008) Using remotely sensed night-time light as a proxy for poverty in Africa. *Population Health Metrics* 6(5): 1–15.
Pfeffer, K., Deurloo, M.C. and Veldhuizen, E.M. (2012) Visualising postcode data for urban analysis and planning: The Amsterdam city monitor. *Area* 44(3): 326–335.
Pfeffer, K., Baud, I., Denis, E., Scott, D. and Sydenstricker-Neto, J. (2013) Participatory spatial knowledge management tools: Empowerment and upscaling or exclusion? Information. *Information, Communication & Society* 16(2): 258–285.
Piotrowski, J. (2014) Big obstacles ahead for big data for development. Posted online 15 April, http://www.scidev.net/global/data/feature/obstacles-big-data-development.html, accessed 12 June 2014.
Press, G. (2013) A very short history of big data. Posted on 13 September, http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/, accessed 24 October 2014.

Rai, S. (2014) India's push for 100 smart cities has tech firms scrambling for contracts. Posted 31 July, http://www.techrepublic.com/article/indias-push-for-100-smart-cities-has-tech-firms-scrambling-for-contracts, accessed 22 October 2014.

Richter, C. (2014) Digital transformations in Indian cities: Between paper list and GIS map. PhD Thesis. Enschede, The Netherlands: University of Twente.

Sevtsuk, A. and Ratti, C. (2010) Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *Journal of Urban Technology* 17(1): 41–60.

Shelton, T., Zook, M. and Wiig, A. (2015) The 'actually existing smart city'. *Cambridge Journal of Regions, Economy and Society*, 8(10): 13–25, available at SSRN:http://ssrn.com/abstract=2477482.

Smith-Clarke, C., Mashhadi, A. and Capra, L. (2014) Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks. In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. (pp. 511–520). April 2014, Toronto, Canada: ACM.

Steenbruggen, J., Borzacchiello, M.T., Nijkamp, P. and Scholten, H. (2013a) Data from telecommunication networks for incident management: An exploratory review on transport safety and security. *Transport Policy* 28(0): 86–102.

Steenbruggen, J., Borzacchiello, M., Nijkamp, P. and Scholten, H. (2013b) Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: A review of applications and opportunities. *Geojournal* 78(2): 223–243.

Sui, D. (2014) Discussant in the panel: alt.conference on Big Data: Lightning Talk Discussion, organized by Jim Thatcher and Andrew Shears, AAG Annual Meeting 2014, Tampa, Florida.

Takhteyev, Y., Gruzd, A. and Wellman, B. (2012) Geography of twitter networks. *Social Networks* 34(1): 73–81.

Tatem, A.J. *et al.* (2014) Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria Journal* 13(52): 1–15.

Taylor, L. (2014) Sustainable data science for sustainable cities: Big data and the challenge of urban development. Opinion paper. Bonn: EADI – chance2sustain.

Taylor, L. (under review) No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *Environment and Planning D*, submitted, at: https://www.academia.edu/7502204/No_place_to_hide_The_ethics_and_analytics_of_tracking_mobility_using_mobile_phone_data.

Taylor, L. and Schröder, R. (2014) Is bigger better? The emergence of big data as a tool for international development policy. *GeoJournal* 1–16, http://link.springer.com/article/10.1007/s10708-014-9603-5.

Taylor, L., Cowls, J., Schroeder, R. and Meyer, E.T. (2014) Big data and positive change in the developing world. *Policy & Internet* 6(4): 418–444.

Thatcher, J. and Shears, A. (2014) AAG Annual Meeting, Tampa, Florida.

Torfing, J., Peters, G., Pierre, J. and Sørensen, E. (2012) *Interactive Governance: Advancing the Paradigm*. Oxford, New York: Oxford University Press.

Townsend, Anthony M. (2013) *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. Cambridge, MA: W. W. Norton & Company, p. 400.

TT Crime (2014) 1994 to present, crime statistics, http://www.ttcrime.com/stats.php, accessed 22 October 2014.

Tufekci, Z. (2014) Engineering the public: Big data, surveillance and computational politics. *First Monday* 19(7).

UN Data (2014) Mobile-cellular telephone subscriptions per 100 inhabitants, http://data.un.org/Data.aspx?q=mobile&d=ITU&f=ind1Code%3aI911, accessed 23 June 2014.

UN Global Pulse (2012) Big data for development: Opportunities & challenges. White paper, Global pulse, New York, May, http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-GlobalPulseMay2012.pdf.

UN Global Pulse (2014) Understanding the post-2015 global conversation through big data, http://www.unglobalpulse.org/projects/Post2015, accessed 20 June 2014.

UN-Habitat (2013) *State of the World's Cities 2012/2013: Prosperity of Cities*. New York: Routledge.

Verrest, H. (2013) Rethinking microentrepreneurship and business development programs: Vulnerability and ambition in low-income urban Caribbean households. *World Development* 47: 58–70.

Wesolowski, A. *et al.* (2012) Quantifying the impact of human mobility on malaria. *Science* 338(6104): 267–270.

Zimmermann, R., Lawes, C. and Svenson, N. (2012) Caribbean Human Development Report 2–12: Human Development and the shift to better citizen security. New York: UNDP.