Article

No place to hide? The ethics and analytics of tracking mobility using mobile phone data

Linnet Taylor
University of Amsterdam, Netherlands

Abstract

This paper examines the ethical and methodological problems with tracking human mobility using data from mobile phones, focusing on research involving low- and middle-income countries. Such datasets are becoming accessible to an increasingly broad community of researchers and data scientists, with a variety of analytical and policy uses proposed. This paper provides an overview of the state of the art in this area of research, then sets out a new analytical framework for such data sources that focuses on three pressing issues: first, interpretation and disciplinary bias; second, the potential risks to data subjects in low- and middle-income countries and possible ethical responses; and third, the likelihood of 'function creep' from benign to less benign uses. Using the case study of a data science challenge involving West African mobile phone data, I argue that human mobility is becoming legible in new, more detailed ways, and that this carries with it the dual risk of rendering certain groups invisible and of misinterpreting what is visible. Thus, this emerging ability to track movement in real time offers both the possibility of improved responses to conflict and forced migration, but also unprecedented power to surveil and control unwanted population movement.

Introduction

Reliable data on human mobility are scarce, and especially so in low- and middle-income countries (LMICs).[1] Mobility data constitute one dimension of the problem of missing statistics outlined by Jerven (2013), where the infrastructure and resources to gather reliable economic and population information in LMICs are often severely limited. A 'datafication' turn[2] is taking place in academic and policy research, enabled by born-digital datasets of unprecedented size[3] ('big data') from new digital technologies.

Corresponding author:
Linnet Taylor, University of Amsterdam, International Development Studies, Nieuwe Achtergracht 166, 1018 WV, Amsterdam, Netherlands.
Email: l.e.m.taylor@uva.nl

Datafication in LMICs is driven at least partly by the rise in mobile phone and internet use (ITU, 2013) worldwide. 'Big' digital data is starting to present an answer to the scarcity of up-to-date, granular information sources about LMICs (Taylor and Schroeder, 2014).

The datafication turn is worth evaluating because it has profound implications for what Scott (1998) has termed the 'legibility' of the subjects of development. Scott's term refers to the ways in which high-modernist, technocratic means are used to shape populations into more governable form. I will argue that the current surge in development research using mobile phone data highlights the gap between legibility and understanding. Separating out often-conflated types of new digital data, this paper focuses on what Hildebrandt (2012) has termed 'observed data' collected by mobile network operators, distinguishing this from volunteered geographical information which raises different issues regarding user awareness and consent. One chief feature of big data is that it is primarily gathered and processed by corporations: for example, Crampton et al. (2014) note that 'three quarters of the imagery utilized by the [US] National Geospatial-Intelligence Agency (NGA) derives from nongovernment or commercial sources'. This suggests that conclusions based on big data primarily represent the perspectives and aims of those with the influence and resources to channel and access it.

There are reasons to be particularly wary of what Pentland (2011) has termed 'the god's eye view' provided by big data when it comes to mobile data about lower income countries. Massey (1993) has argued that who one is determines how one may move, and that the new technologies of mobility do not benefit everyone equally. If we apply her logic to mobility as reconstructed via mobile phone traces, we see that they may expose some people and hide others, and that a discourse about the universality of these signals masks their unevenness and complications. We should also ask how using such remotely collected and processed data affects our ability to research mobility, and whether it may deepen the divide between researchers and subjects while also increasing the potential for powerful interests to monitor and intervene with those subjects. Asking this question takes up the challenge issued by Dalton and Thatcher (2014) for a 'critical data studies' which treats big data as a product of, and a contributor to, power geometries of place and space. Building on their critique, I explore how research using mobile phone data gives rise to a particular kind of 'socially produced' space by reconstructing and visualising people's activities and movements, a space that is at once virtual and material (Crampton et al., 2013). The virtual character of such space necessitates interpretation, but its material aspect lends itself strongly to intervention by those interested in monitoring, governing and controlling.

Given the new potential for identification and intervention that comes with 'seeing' through digital traces (this term is used here to denote the diverse signals produced through the use of digital technologies and applications (Thatcher, 2014)), this type of research may have serious implications for its subjects. Crampton et al. (2013: 138) warn that big data has institutional politics and that it will therefore tend to be operationalised towards 'particular, and not always benign, ends'. In the same vein, Gabrys points out (2014) how 'environmental technologies' such as mobile phones, which tell a story of people's activities, may channel and shape the way people relate to authorities. This paper will explore how research using mobile data may empower the remote observer, and the ethical and practical implications of that empowerment.

I will look at two categories of problem. First the risk of misuse, given that the increase in availability of data on people's locations and communication networks has occurred relatively fast – over less than a decade – and is not yet balanced by the emergence of privacy legislation or ethical frameworks for the use of data. In particular, the use of big

data brings up the novel problem of *group privacy*, where the harm occurring from data misuse takes place on the aggregate level rather than the individual, and which stretches current conceptions of privacy. Second, there is the risk of misunderstanding the data, since those who can access it may not be those with a contextual perspective. Due to issues of access (since mobile data are proprietary to the network providers who collect them), and capacity (the resources, computational power and technical skills to analyse them), mobile data have been primarily taken up by the international data science community rather than by social scientists. Research on mobile phone traces therefore tends to be conducted remotely by non-social scientists, which inevitably changes the relationship between the researcher, the research institution and the data subject and often makes understanding the local context problematic (a problem identified by Burns (2014) in the humanitarian field).

The paper is set out as follows: the next section provides an overview of how mobile data research has evolved, with particular attention to sub-Saharan Africa where these data provide a potential step change in terms of detailed mapping. I then present a case study of the first instance where data from an LMIC (Côte d'Ivoire) were released to researchers on a large scale (Netmob, 2013) and highlight the ethical and methodological questions raised by the project. I first explore the potential for technical problems of misinterpretation, and next the particular risks mobile data imply for data subjects in lower income locations, and the new ethical questions they pose. I conclude by drawing these problems together around the notion of 'function creep', or how data 'collected and used for one purpose and to fulfil one function, often migrate to other ones' (Lyon, 2008: 6, following Winner (1977)). I argue that it is time to urgently review and evaluate the use of mobile phone data from low-income countries for tracking mobility and to develop a new ethical and regulatory framework for research and analysis.

A note on methodology

This paper is based on a series of 60 interviews on the use of big data in development policy and planning, along with two years' background research on the interface between data science and international development policy. A number of international conferences, workshops and public discussions hosted by institutional actors in the field were attended over the two-year period, and knowledge was also gathered from mailing lists and other online discussions. The interviews on big data in the field of international development were conducted with academic researchers and private-sector data scientists working with big data on questions relevant to LMICs, including 10 with researchers or managers directly connected to the Data for Development project run by Orange. These were conducted during the months following the 2013 Netmob conference in Boston, with follow-up research conducted over the ensuing two years. Interviewees were selected using a purposive sampling process focused both on the most relevant projects and on those with an overview of the state of the art in big data research. Interviewees are identified in this paper except where they requested not to be.

Background: Mobility measurement and the emergence of big data sources in LMICs

It is a persistent challenge for states to track people's international movement with any degree of accuracy (Makaryan, 2012), particularly in regions where informal migration dominates. Sub-Saharan Africa, the location of the case study for this paper, is just one region where large numbers are believed to be missing from official statistics (Carr-Hill,

2013). However, the African continent, and LMICs in general, are seeing a trend towards greater digital visibility (Taylor and Broeders, 2015). There has been an exponential rise in mobile phone use in sub-Saharan African countries in particular, which still constitute the world's most technologically marginalised region: the proportion of people on the continent with a mobile subscription rose from 12.4 to 63.5% from 2005 to 2012 (ITU, 2013). Citizens of LMICs are expected to provide the majority of geocoded digital data by 2020 (Manyika et al., 2011). As people use mobile phones more, however, they are also becoming more identifiable and visible. In 48 out of 55 countries on the African continent, compulsory SIM card registration has been adopted (Donovan and Martin, 2014), so that any activated SIM card is now linked to a recorded individual identity. Smartphones, which are more likely to identify their user through app use and more frequent contact with the network, are becoming more accessible: the share of sub-Saharan Africans with smartphones is the world's lowest at 10.9% (Donovan and Martin, 2014), but has risen six-fold since figures became available in 2010.

Research on human mobility using mobile phone traces is a recent phenomenon (for an overview see Blumenstock (2012)) and has expanded since around 2006. Researchers began by uploading tracking software onto consenting subjects' phones (Eagle and Pentland, 2006), but soon gained access to data directly from mobile network providers, leading to larger scale research and greater analytical power (e.g. Gonzalez et al., 2008). One of the first research teams to apply this methodology to mobility in an (unnamed) developing country was that of Eagle et al. (2009), who compared rural to urban mobility. They were followed by Soto et al. (2011) who compared mobility to income levels in Latin America, and Pindolia et al. (2012), who explored linking and merging other datasets to identify data subjects' demographic and socioeconomic characteristics.

This evolution from simply tracking movement to a more behavioural perspective has led to the application of a 'development' lens to mobile data from LMICs. For example, Blumenstock (2012) conducted a study of 1.5 million mobile phone users in Rwanda, with the stated aim of 'demonstrat[ing] how these data can be used to improve development policy' (2012: 121).

The growing troves of location data for areas that are still quite sparsely mapped due to poverty have led to excitement in the development policy community. The mobile phone industry association, the GSMA, has developed a 'Mobile for Development Intelligence' programme[4] with the aim of persuading mobile providers to share data with researchers and development organisations, with the rationale that 'open access to high quality data will improve business decision making, increase total investment from both the commercial mobile industry and the development sector and accelerate economic, environmental and social impact from mobile solutions'. The director of the UN's data-for-development initiative, Global Pulse, has stated that it is now possible to 'establish a 360-degree observatory on poverty' using as many sources as possible, including mobile data (Kirkpatrick, 2013) – an image which initially corresponds uncomfortably with that of Foucault's Panopticon (1977). In the Panopticon, an ideal prison designed by the 19th-century reformer Jeremy Bentham, every prisoner is visible at all times to the jailer, and must therefore behave as if they are being watched, although they cannot know at any given moment whether anyone is looking. The developing-world Panopticon, however, is possibly more insidious because it lacks such a mutual awareness between subject and observer, since in the regions where mobile phones are a new technology subjects are unlikely to be aware they are being observed, or by whom.

As mobile data become more available, the risk grows that the increase in granularity they offer will be used to inform new claims to possession. Writing of the interplay between

cartography and development, Bryan (2011: 42) notes that maps are 'a technique of calculation that are used to calculate distributions, organize markets, and identify territories and populations, and [are] associated with notions of government as attaining the 'right disposition of things''. Mapping tends to precede possession: Harley has highlighted how the 'political unconscious of the map' (1989: 528) inevitably tends towards domination, showing how the maps of the Conquista informed an 'anticipatory geography [which] served to frame colonial territories in the minds of statesmen and territorial speculators back in Europe' (1989: 532). Just as paper maps did for the Conquista, the evolution of real-time, remote mapping of people's mobility may also enable evolutions in remote colonialism. Kirsch's work (2014), for example, examines how mapping the Philippines enabled the US to claim them as a territory without annexing them as a colony: it owned the information, so it need not occupy the space itself. Owning the map enabled the dominant power to demand taxes, conscript labour and determine who should share the profits. Conversely, this also meant that those who were left off the map then struggled for representation in the new political order. The gaps and uncertainties left by today's big data are similarly important in understanding how it represents, or fails to represent, data subjects. For instance, Stephens (2013) highlights how the uneven adoption of digital technologies can produce highly biased accounts of space, and Graham (2010) has explored the gaps and invisibilities that maps derived from digital traces leave. Furthermore, Kingsbury and Jones (2009) and Zook and Graham (2007) among others have highlighted the uncertainties that exist, or can be inserted into, apparently data-rich constructions of space. Their analyses of the chaotic and qualitative nature of volunteered digital geo-data demonstrate how big data can be simultaneously discursively powerful and analytically unreliable.

The new mappings of mobility made possible by mobile data also create a range of possible futures. Crampton (2011) encourages us to ask what kind of territories are produced by analyses of new types of data such as mobile phone traces, and how these may alter understandings of space. These data cross national and regional boundaries, since mobile phone services are designed to 'roam' across borders for both user convenience and corporate profit, and therefore produce not the (over)-clear bounded and contained spatial definitions of GIS but 'fleeting, transactional records' (Crampton, 2011: 4) of activities rather than places. Big data enables us to map people and movement without necessarily mapping land. The people are the territory. Thus, if the aim of much GIS work has historically been to establish claims to land and to govern people, the new data technologies are in comparison more remote: they allow the viewer to track, often in real time, and to influence. Particularly with regard to the transgression of state boundaries involved in irregular migration, as will be explored here, the new data from digital traces lend themselves to a post-Westphalian politics of influence and indirect action.

Case study: The 'Data for Development' challenge

The 'Data for Development' (D4D) challenge, organised in 2012–13 by the mobile provider Orange, illustrates the central issues of ethics and interpretation stemming from mobile data research on LMICs. It involved the first major release of mobile phone calling records from an African country for research purposes and was designed to encourage both basic and applied research using an anonymised dataset of a year's call records from all Orange's subscribers in Côte d'Ivoire. The aim of the challenge was to 'help address the questions regarding development in novel ways' (Blondel et al., 2012). The company benefited from the challenge in terms of positive publicity in its main (European) markets, although it was not

widely publicised in Côte d'Ivoire. Of the 74 papers presented at the resulting conference, two-thirds dealt with tracking mobility. Another section dealt with social and economic development (mapping poverty, tracking economic and social activity), a third with data mining and a fourth with health and methods for tracking epidemics (Netmob, 2013). As well as being the first release of mobile data from an African country to the general research community, it was also the first to be labelled a 'development' project, and gained huge publicity after it was endorsed by the United Nations, the World Economic Forum and a host of high-profile academic institutions including MIT and Cambridge University (Netmob, 2013).

The dataset, comprising data from around five million users, consisted of four elements: records of 2.5 billion within-network calls and SMS exchanges over the period of a year, the spatial trajectories of 50,000 users with high resolution over a period of two weeks (based on calls and SMS traffic), the trajectories of 500,000 users at lower resolution over the course of the year and communication subgraphs showing the communication networks of 5000 users over the year. The dataset was released through a formal application process to 150 teams of researchers worldwide, mainly from the disciplines of mathematics, physics and computer science.

The project highlights several of the issues which arise around the release of entirely new types of data in the development field: the lack of enforceable ethical and legal parameters for the sharing and reuse of data, practical issues to do with anonymisation and the problem of 'ground truth' (Pickles, 1995). First, Orange identified the lack of a locally binding legal framework to govern the release of the data, and therefore had to release it within an ad hoc framework where the only binding commitment to privacy and data protection was made between researchers and the company through a nondisclosure agreement. A company vice president explained the problem: the other countries of francophone West Africa use a version of France's stringent data protection regulations established by the 'Commission Nationale Information et Liberté', but Côte d'Ivoire is not a signatory. The company therefore referred to its operating license with the Ivoirian government's communications ministry, which allowed data reuse for 'research or artistic purposes'.

> So with that in mind, we said okay, we are now going to do something that will be "best in class" anonymisation and management process, that will be going further than the most stringent law on the planet would require us to do for a research challenge project. And so that is what we did. (Nicolas de Cordes, Vice President, Marketing Vision, Orange-France Telecom Group, interviewed 16 April 2013)

Having checked the rules in its contract the company then made the Ivorian Communications minister aware that they would be making the release. The firm, however, was not legally constrained to any particular course of action other than the very broad definition of permissible data reuse contained in its operating license, and the ethical constraints were thus defined by industry standards alone:

> we had a discussion with our colleagues, the marketing department of Orange in Ivory Coast, and they said that there were barely any constraints from Ivory Coast's regulator on privacy, and so after I explained that our intention was to be very strict on these aspects, they said it is not necessary to ask, you can move on, so we moved on, we applied as I said the most stringent rules for an organisation. (Nicolas de Cordes)

There is no doubt that Orange made its preparations in good faith, but it is clear that even applying 'the most stringent rules' nevertheless constitutes self-regulation. To check it was not infringing customers' privacy, the company also asked a team of researchers to 'attack' the dataset (try to de-anonymise it to determine threats to user privacy) before it was

released (Sharad and Danezis, 2013). However, this uncovers the second problem particular to large data releases, anonymisation, and illustrates why it is different in a developing country.

Sharad and Danezis observe that each of the four datasets released as part of the Orange challenge makes it possible both to build a picture of (anonymous) individuals' social connections and to connect each individual with a geographic area of origin and then track their movements. They then note that a common way to de-anonymise individuals is to merge and link these network-based identifiers with others gleaned from online data from the area in question. However, as they point out, there is little online activity in Côte d'Ivoire, and other datasets that might help to de-anonymise individuals in this way are not available. Instead, they posit a situation where someone living in a small village is using a mobile phone intensively. Local residents are likely to be able to identify this person, and in turn, 'if this individual can be persuaded sell his call history then this can be combined with the dataset to mount very potent attacks which can de-anonymise a large number of people and cause significant breach of privacy' (Sharad and Danezis, 2013: 12). Thus, instead of employing complex techniques to identify individuals remotely using online data, one can simply ask the person who makes the most calls in their village to identify first one user, and from there de-anonymise large portions of the dataset. This problem, as the authors point out, is particular to areas which are sparsely populated such as villages and rural hamlets, where users stand out in a way they would not in a country with greater mobile penetration and more wealth and infrastructure. Again, significantly, the challenge is to the anonymity of the group as a whole, rather than only the individual who is persuaded to sell his or her calling history.

The third issue Orange's data release illustrates is that of ground truth: the tension between data scientific skills and contextual understanding. In the present example, the field of international development has been characterised by a huge range of understandings of what constitutes 'development' (e.g. Chambers, 1990; Collier and Dollar, 2001; Lucas, 1988) and contestation of these understandings (e.g. Escobar, 2011; Sachs, 2005; Sen, 1999). The D4D challenge entered this fray with what de Cordes described as 'a fuzzy objective' in terms of its definition of development, and with what Vincent Blondel, organiser of the challenge, described as the desire to do something 'big and interesting' (Vincent Blondel, Professor of applied mathematics at the Université Catholique de Louvain, interviewed 29 March 2013). The dataset was released to the research community in general, but could only be analysed by those with the proper 'big data' technical skills of advanced mathematics and statistics. This created a divide between those who could analyse the data and those who could understand its context. A commercial transport researcher who headed one of the research teams was aware of this problem:

> Of course, this was the most interesting phenomenon of all – that we were just sitting here in the Netherlands getting this data, taking a network from the internet, getting all the other data from the internet, then, of course, the strange thing – going to a conference where nobody from Ivory Coast is, and we're telling them, "Here's your transport model". (Peter van der Mede, Director, Goudappel Coffeng, interviewed 2 May 2013)

Very few of those involved in the D4D challenge had direct experience of Côte d'Ivoire. The data were collected after a year of civil war (2011) where ethnic tensions had been extreme, the government had fallen and been replaced by another, and conflict had profoundly influenced Ivorian mobility and activities. However, these conditions were not reflected in the information offered to researchers to accompany the datasets. Van der Mede adds: '. . . a lot of people were just doing, I think, the things they are always doing in their research area and just using another data set to do the same thing'. The organising team

included a professor from Bouaké as a consultant to the project, but of the 150 research teams which received and analysed the data, one was based in Africa (a team from Cameroon), and only one of the other researchers said that they had visited Côte d'Ivoire and conducted interviews to understand the data's potential biases and limitations. There were thus limited opportunities for researchers to understand or be influenced by local understandings of what might constitute development or privacy, or to be sensitive to what kinds of research ethics might be appropriate for a fragile, post-conflict state.

The D4D project in 2013 represented a breakthrough in the willingness of network providers to release mobile calling data beyond the context of humanitarian emergencies, something they have so far been reluctant to do. The resulting papers represent both the state of the art in terms of mobile data analysis with relation to a lower income country, but also the problems that context raises: the need to piece together an ad hoc ethical framework within which to make data available to researchers, since no review board governs research in the corporate sphere; the need for corporations to self-regulate with regard to sharing users' data from countries without strong data protection standards; the fact that today's anonymisation may be tomorrow's identification, and the lack of contextual understanding on the part of 'big data' researchers from disciplines outside the social sciences. The D4D release also suggests that the problem of identifying individuals (through 'personally identifiable information', according to the industry terminology) may become dwarfed by the problem of identifying groups through 'ontologically constitutive information' (Floridi, 2013) – whether cohorts, areas or communication networks. In a situation of conflict where group allegiances have heightened implications, the ability to identify a group's location, leaders, boundaries and communication networks may provide a means for hostile state or local forces to threaten the safety of those identified. The next sections discuss the problems of data protection, ethics, risk to data subjects and problems of interpretation raised by this type of research project in more detail.

Mobile data's potential for misuse and inaccuracy: Issues of interpretation and bias

For a researcher trying to determine a subject's location using mobile data, two levels of specificity are possible depending on the type of data available. First, calling records show which of a network's antennae is being used, and thus the user's movement from the vicinity of one antenna to another. Alternatively, more specific location data can be gathered as a phone's SIM card automatically checks in with its nearest antenna. These data are particularly detailed in the case of smartphones, which constantly check for updates to email or other applications. As noted by Michael and Clarke (2013), 'mobile devices auto-report their presence 10 times per second'. These real-time data have until recently been accessible only to mobile providers themselves, governments and law enforcement, and (in the case of Apple devices) advertisers (Michael and Clarke, 2013), but not to researchers or humanitarian organisations. This has started to change, however, since the Haiti earthquake provided a strong case for the use of this type of continuous location data for humanitarian purposes, namely providing epidemiological data on the cholera outbreak that followed the earthquake (Bengtsson et al., 2011).

Using geocoded mobile data as a proxy for human mobility involves a two-way translation process. First, meaning must be ascribed to particular patterns and signals. Second, meaning must (usually) be subtracted from the dataset through anonymisation, using techniques that still leave enough patterns and signals to read the data with some degree of specificity. The process of using this type of data, then, is a balancing act between uncovering and obscuring specificity. Each of these two conflicting processes, however, has

its problems: the data may be non-specific in ways the researcher does not understand due to cultural or geographic distance, and the necessary qualitative information is not easily accessible to researchers who are not social scientists.

One of the problems for those trying to see movement clearly using this type of data is *numerical accuracy* regarding data subjects. On average, the presence of antennas will correlate fairly closely with the presence of people: a mobile provider's ability to collect location data on its users is dependent on their connecting with antennas, and since remote locations have fewer antennas (de Montjoye et al., 2013), where the population is less dense, fewer signals will be collected. So for example there will be both more signals and more people in urban areas than in rural ones. This correspondence between signals and individuals becomes unreliable, however, when large numbers of people move through remote areas with few antennas. This can occur under normal conditions when people go to remote rural locations for religious or other cultural gatherings (e.g. prayer camps, initiation rituals, pilgrimages), but this inversion of the ratio of people to signals is particularly likely under conditions of duress and forced movement, such as occurring during emergencies. Then, large populations may become present in remote areas, either fleeing towards international borders (as in the Darfur crisis) or escaping violence by fleeing into the forest (as during the conflicts in the Democratic Republic of Congo or Rwanda). Thus, people's visibility can be in inverse relation to their security.

Another dimension of numerical accuracy is the problem of *multiple identities* within the dataset: it is risky to assume that one signal represents one individual, since in many places a lack of coverage can result in people having multiple SIM cards from different providers to get the best chance of a signal in remote locations (Bengtsson et al., 2011). Equally, one SIM card can have multiple users. This can mean that if using data from multiple providers, several different profiles may represent a single user, or the reverse.

Another category of problem in studying movement through remote or low-income areas is *locational accuracy*: given that the locations of infrequent callers are updated less often than those of frequent users (Bengtsson et al., 2011), those with less resources to buy calling credit, or to charge their phone's battery, are less visible. This is a particular problem in the case of forced movement where people may become unable to recharge their phones as they move. A related problem can be identified where a phone user may run out of credit or battery power while moving, and effectively be pinned to the map at the last place their phone made contact with an antenna. Thus, people may be only fuzzily visible or may be first visible and then invisible – demanding that the researchers come up with a way of dealing with missing quantitative data in a context that calls for purely qualitative information.

Finally, the data's accuracy also depends on the researcher's *degree of access*. Calling records are vastly less reliable than continuous SIM (or even phone) location details. If all that can be accessed is calling records, accuracy in tracking movement is vastly reduced. If, however, there is access to the location data being sent by the phone itself, the potential to track people is limited only by the phone's access to the network – and by the researcher's ability to understand how the data may represent what is happening on the ground. This last problem is the largest: the dilemmas of data interpretation are dwarfed by the challenge of understanding ground truth, and in turn this challenge is intensified by the cultural and experiential gap between researcher and data subject, as exemplified by Orange's D4D project.

Risks, rules and remedies

Mobile traces are currently used by researchers for purposes that span very different timescales and justifications, ranging from real-time humanitarian response

(e.g. Bengtsson et al., 2011) to understanding socioeconomic change (Soto et al., 2011). Nevertheless, there are many commonalities in the risks such research may pose to its subjects. Digital data have a tendency to multiply and replicate: once released, it is impossible to put it back in its box. Blumenstock (2012: 121) writes of mobile phone data research that 'there are no pragmatic recipes for how to deal with what is an inherently ethical dilemma, [and researchers] cannot reject the possibility that derivative methods would be used for less desirable purposes'.

For LMIC citizens, leakage and reuse of their data tend to occur without accountability because a lack of enforceable regulation makes subsequent uses of data beyond the original research harder to track and control. The US has minimal data protection standards which only apply to citizens, Canada similarly lacks strong regulation and the EU's data protection directive (95/46/EC), while much stronger than either the US or Canada's, only covers EU citizens or uses of data within the EU's territory (Gasson et al., 2011). Meanwhile, a patchwork of data protection laws exists across the African continent (Greenleaf, 2013: 11) but few are enforceable so far, nor do governments always have the capacity to audit what multinational corporations are doing with their citizens' data. This means that in most cases, as with the Orange D4D challenge, currently only self-regulation by corporations and independent researchers stops these data from becoming a resource for surveillance and control or profiling and discrimination. Orange recognised this problem and incorporated an ethical review process into its second data challenge (Orange, 2015) to consider each submission for its potential risks to data subjects on an individual and collective basis. The review process constituted a rare interface between positivist data science and a more situated and nuanced social scientific understanding of risk and ethics in research since, so far, the ethics of GIS as discussed by Schuurman (2000) or Sieber (2006) have not been part of the discussion about the uses of mobile data in LMICs. Because the data are collected and shared by corporations, ethical decisions in this area are perceived by researchers as being taken upstream of their involvement with the data.

An ethical committee such as Orange's is faced with two important issues. First, the growing mismatch between 'personal data' and sensitive data, and second, the ethical implications of the choice to make people visible in new ways and to new observers. With regard to the first, big data is changing the definition of personal data. For example, although privacy rules around mobile data stress the 'proper anonymisation' of 'personally identifiable information' (Global Pulse, 2013; GSMA, 2011), large-scale mobile datasets may expose personal characteristics despite anonymisation either because they can be re-identified, or because the risk occurs on an aggregate level. Beyond the possibility of re-identifying individuals in a dataset, there is also the problem of group visibility. This presents a paradox in terms of current anonymisation standards. In practical terms, researchers use two main approaches to privacy concerns with geocoded data: first, various methods to blur or restrict the data available on a particular query (k-anonymity), and second, introducing tracking uncertainty by pruning data in order to reduce 'time to confusion' – the length of time an adversary can accurately track an individual (see Manohan (2009) for an overview). This means that methodologically, both these methods rely on altering data in order to make the group visible but not the individual.

However, where mobile phone data make visible the movements or network structure of a group, problems arise since its members may not need to be identified in order to be subjected to harm. Sharad and Danezis (2013) observe that even without identifying information it is possible to track the movement of groups if they have particular calling patterns or a particularly active leader, and Blumenstock (2012) shows that if a large,

anonymised mobile dataset can be combined with a much smaller survey-based study of individuals within that dataset, it becomes possible to effectively de-anonymise the larger dataset in terms of group characteristics, i.e. to identify groups by age, gender, profession and employment status and to thus derive group-specific findings. De Montjoye et al. (2013) have shown that mobile data can be an extraordinarily efficient way of identifying groups and social networks, and identifying when they move simultaneously. Furthermore, advances in agent-based modelling (e.g. Kniveton et al., 2011) suggest that mobile phone data may soon enable researchers to predict, rather than just observe, human mobility – raising a host of new ethical concerns with regard to pre-emptive responses on the part of national and international authorities.

This brings us to the second issue, namely ethics. Some attention has been paid to the implications of knowing the movements or activities of groups such as dissident networks (MacKinnon, 2012; Morozov, 2012), and this research may need to be scaled up to address the ethics of research using mobile data from LMICs which may have particular problems with political or ethnic violence and discrimination. Broeders (2011) has outlined the conditions under which group mobility may be tracked, including political emergencies, environmental crises and conflict situations. One example of the problems mobile data may raise in this context is provided by the dataset from Senegal released in Orange's second D4D challenge, which relates to the period of the 2014–15 Ebola epidemic in West Africa. Some researchers proposed to use the data to derive models for quarantine policy, based on group-level movement and social network dynamics, which were judged by the ethical committee to have significant potential for misuse.[5]

If data pose a risk to groups as much as to individuals, this raises new questions about informed consent. Bernal (2010) argues that since personal data are constantly updated, a system of real-time 'collaborative consent' must be developed. However, this can work only where users are continually connected, literate and aware of the problems of privacy that their technology use may pose. The only apparent alternative, which Solove (2013: 1881) characterises as a 'paternalistic' approach, requires legal and rights-based standards for privacy and data protection, plus the effective rule of law to ensure enforcement. This leaves many researchers working with mobile data from LMICs without an ethical framework to guide their decisions.

The difficulty of anonymising mobile data also raises the practical problem that while the new mobile datasets offer the possibility of observing human mobility (and other activities) in unprecedented real-time detail, conducting ethical research involves the effort *not* to see too clearly – i.e. to reduce the possibility of identifiability and potential harm to data subjects. Faced with this dilemma, researchers must make explicit choices about what to see, and what to make visible to others. This means balancing the risk of being identified with that of *not* being identified, since as well as potentially increasing vulnerability, mobile traces may also constitute a way for those in need of visibility to become visible – Parr and Fyfe's 'right to be seen' (2012) – and better protected from danger, deprivation or exclusion. An interviewee for this research who was involved in a project to share data for humanitarian purposes advocates a utilitarian perspective:

> You always need to make this trade-off between the potential good consequences of analysing operator data and the general rule of not accessing other people's personal data without their consent. In some situations it would be a violation of people's rights if someone did not analyse their data – if that would prevent them from, for example, receiving critical emergency relief. The complicated and important discussion is in what type of situations, for what purposes, by whom, and in what format operator data should be used. (Anonymous – interviewed 17 June 13)

Thus along with the risk of becoming visible goes the risk of invisibility, and of the consequent inability to claim one's rights and entitlements.

The risk of profiling and harm through the identification of individual or collective data subjects in LMICs, and the lack of viable models for mitigating it, highlights a larger research gap regarding data sharing in lower income countries. Research on when people consider mobile tracking unjustified has been confined to industrialised countries and higher income populations (e.g. Barkhuus and Dey, 2003), as has research on the ethics of tracking (e.g. Michael et al., 2006). However, where a state is fragile, post-conflict or otherwise politically unstable, the context is arguably very different. The potential dangers of making data subjects more visible are greater, since in less politically stable environments there are fewer limitations on the ability to profile negatively or harm them. Since the global centres of data science are located in high-income countries, LMIC data subjects are also less likely to discover that their data are being used or be able to contest its use. Data scientists, too, are generally ill-equipped to deal with the ethical issues that arise in their research. Those who analyse big data come mainly from disciplines such as theoretical physics, computer science and mathematics and have little or no training in how to approach complex and socially situated questions of power asymmetries and the rights of data subjects with regard to visibility or invisibility. For this reason, new ethical codes and interdisciplinary collaboration are becoming essential for big-data research focusing on development and crisis response.

Function creep: From humanitarian to 'development' uses of mobile data

The risks of data misuse in the context of mobile traces mainly stem from power and knowledge asymmetries between researchers and data subjects and between corporations and states. The sharing and reuse of mobile data are potentially powerful ways of reproducing these asymmetries, particularly under conditions of data maximisation and minimal regulation, and even more so where such data maximisation is labelled as being in the cause of development (Taylor and Broeders, 2015). Under these conditions, what Lyon (2008) has referred to as *function creep* is inevitable. Lyon identifies all surveillance as occurring on a spectrum ranging from care to control: function creep is a shift from care towards control. The use of mobile location data has already shown function creep: one of the earliest applications of such data (Arai, 2006) was for 'the security of children and elderly people' – both vulnerable groups where informed consent was less easy to obtain. Similarly, in the US such data were at first only available to emergency services in case of personal danger. Soon, however, access was also extended to law enforcement and other authorities (Pell and Soghoian, 2012) and in 2013 the revelations of Edward Snowden made it clear that many governments were collaborating to use mobile data for mass surveillance.

The D4D challenge, along with other research such as that of Pindolia et al. (2012) shows how mobile data research has evolved to a point where it can serve international development and migration policy concerns, including the prediction, planning or prevention of mobility. In contrast to humanitarian organisations oriented towards care – responding to mobility in cases of natural disasters or conflict – governments have an interest in controlling mobility, and specifically in predicting, tracking and preventing unauthorised migration flows towards their borders. Mobile phone data used either as real-time surveillance data or in agent-based models clearly have the potential to help governments pre-empt undocumented migration, thus moving along the spectrum from care to control. One central characteristic of function creep is that it is incremental, which makes it hard to

pinpoint a clear tipping point between acceptable and unacceptable uses of people's data. It is impossible to argue with sharing people's location data in order to protect them from disease, natural disaster or conflict. Nor it is possible to identify a consent problem, since in cases such as these sharing the data is manifestly beneficial to the data subject. However, where on the research spectrum between this kind of immediate threat and local transport optimisation does the data subject's consent become necessary? And how, geopolitically, should a line be drawn between lesser and greater potential harms and more or less trustworthy actors who might gain access to such data? These questions are so far unanswered, as is the larger regulatory question of who is qualified to draw such lines and how they might be enforced. Linus Bengtsson, founder of Flowminder, an organisation that uses mobile data for epidemiology, suggests that the standards used and model for consent are contingent on the circumstances:

> I think somebody needs to make this call, and it will not always be possible to make an informed consent form for everybody, but it really depends on the purpose of what you're doing. (Linus Bengtsson, Flowminder, interviewed 16 May 13)

When balancing contending desires for information in the context of poverty, emergencies and political fragility, it is hard to know who might constitute an honest broker – but it is becoming clear that one will soon be needed.

As the potential of mobile data research edges from care towards control, it also becomes possible to sort and categorise people remotely – a process Scott (1998) has termed making people 'legible'. Legibility, however, does not ensure accountability, especially when it is created through remote data analysis, and may actually leave people invisible in terms of agency and rights. There is nothing participatory or voluntary about mobile data analytics: they tend to lead to mappings more like the colonial cartographies described by Harley (1989) and Kirsch (2014) than those of participatory development projects that incorporate volunteered geographic information. Big data make it almost impossible to be illegible now that even the poorest often have some access to a mobile phone: anyone with a mobile phone makes themselves and their activities legible in ways that cannot be erased. This legibility without accountability is a particular risk in areas where technology corporations such as telecommunications giants are closely allied with the state, as is often true in authoritarian LMICs.

Conclusions

Mobile traces are an important new resource in tracking human mobility, but there is a tension between using these data as an engineering tool for policy-relevant research and understanding its contextual, ethical and political dimensions. The concept of 'proper anonymisation' is not a sufficient response to the complex challenges of using mobile data about potentially vulnerable populations in areas of poverty, political instability or crisis, where risks stemming from privacy violations may be collective as much as individual, and where those risks may involve physical danger rather than the unwanted marketing or identity theft faced by data subjects in high-income countries.

These risks are exacerbated by the likelihood that mobile location data may be misinterpreted by researchers. Despite the deceptively simple correspondence between the movement of a device's signal and that of its user, tracing the movement of people through mobile phone signals requires a process of translation that takes account of technical and social complexity, and that is sensitive to contextual, ethical and political factors. Researchers working remotely and trained in the natural or computer sciences are

generally excluded from the flows of contextual and qualitative information that constitute ground truth for big data and are not connected with data subjects by a tradition of field research. Unless certain gaps can be bridged – between local and international, data science and development research, mobile phone user and researcher – these data may obscure more than they reveal. These problems of heightened risk and restricted information are complemented by a lack of up-to-date, enforceable standards for data protection in the majority of LMICs, creating a potential perfect storm if data are used irresponsibly.

Rather than a single large-scale data disaster, these risks are taking the form of function creep in the use of observed data from LMICs. A growing tendency towards data maximisation, an absence of regulation for powerful corporate and institutional actors, and the shibboleth of 'development' which justifies surveillance under the rubric of care and prevention (Hosein and Nyst, 2013), suggests that the growing legibility brought by mobile data will lead to increased control rather than better understanding and benefits to data subjects. Projects such as Orange's D4D suggest that mobile phone traces may place those in the lowest income countries more on the receiving end of space–time compression than they have ever been, and that debates about the right to move need to take the broader implications of sensing technologies into account. As real-time and predictive analysis of migration becomes more possible, only the lack of an international legal framework for sharing such data keeps such tools from joining the policymaker's arsenal, already used to control who may move and where.

This paper has focused not so much on whether it *should* be possible to trace people through their communications devices, but on *whether* and *how* it is possible to balance the right to privacy with the right to be protected from harm, and the 'right to be left alone' (Warren and Brandeis, 1890) with the right to be seen. Addressing these questions requires interdisciplinary research that must involve a variety of stakeholders: technology firms, country authorities, privacy specialists and researchers with developing-country knowledge. The current system of self-regulation is insufficient given the high stakes involved in mobile-data research on LMIC populations: once data have been shared, it will only replicate.

The ongoing use and sharing of geocoded mobile data therefore presents an argument for some form of ongoing arbitration. As access to data increases, the potential for ethical conflicts grows: people's right to freedom of movement and the freedom to escape danger may clash with concerns over disease transmission, overcrowding or strain on receiving areas. In a best-case scenario, the god's eye view of human behaviour with regard to mobility may enable a timely and appropriate response from authorities without harming data subjects. In a worst-case scenario, however, restriction or harm may be the result. In order to achieve a balance between the right to invisibility and the right to be seen, a better system than self-regulation may need to be established for both corporations and researchers.

Declaration of Conflicting Interest

Funding

Notes

1. I use the World Bank's definitions for grouping countries, see http://data.worldbank.org/about/country-classifications.
2. Mayer-Schönberger and Cukier (2013) use the term to denote the increasing production of machine-readable data that lend themselves to large-scale analytics.
3. This definition of big data is drawn from the work of Taylor and Schroeder (2014) and of born-digital data from Borgman (2014).
4. https://mobiledevelopmentintelligence.com/.
5. Personal communication with company executive from Orange.

References

Arai Y (2006) Geolocation technology and local information in mobile telephony. *Networks and Communication Studies* 20(1–2): 9–25.

Barkhuus L and Dey AK (2003) Location-based services for mobile telephony: A study of users' privacy concerns. *INTERACT* 3: 702–712.

Bengtsson L, Lu X, Thorson A, et al. (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Med* 88: e1001083.

Bernal P (2010) Collaborative consent: Harnessing the strengths of the Internet for consent in the online environment. *International Review of Law, Computers and Technology* 24(3): 287–297.

Blondel VD, Esch M, Chan C, Clérot F, Deville P, Huens E and Ziemlicki C. (2012). Data for development: The d4d challenge on mobile phone data. arXiv preprint arXiv:1210.0137.

Blumenstock JE (2012) Inferring patterns of internal migration from mobile phone call records: Evidence from Rwanda. *Information Technology for Development* 18(2): 107–125.

Borgman C (2014) *Big Data, Little Data and Beyond*. Cambridge: MIT Press.

Broeders D (2011) *SR16: New Technologies and the Monitoring and Management of Migration Flows and Population Displacement*. London: UK Government Foresight Project, Migration and Global Environmental Change.

Bryan J (2011) Walking the line: Participatory mapping, indigenous rights, and neoliberalism. *Geoforum* 42(1): 40–50.

Burns R (2014) Rethinking big data in digital humanitarianism: Practices, epistemologies, and social relations. *GeoJournal* 80: 477–490.

Carr-Hill R (2013) Missing millions and measuring development progress. *World Development* 46: 30–44.

Chambers R (1990) *Rural Development: Putting the Last First*. New York: Longman.

Collier P and Dollar D (2001) Can the world cut poverty in half? How policy reform and effective aid can meet international development goals. *World Development* 29(11): 1787–1802.

Crampton JW (2011) Cartographic calculations of territory. *Progress in Human Geography* 35(1): 92–103.

Crampton JW, Graham M, Poorthuis A, 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.

Crampton JW, Roberts SM and Poorthuis A (2014) The new political economy of geographical intelligence. *Annals of the Association of American Geographers* 104(1): 196–214.

Dalton C and Thatcher J (2014) What does a critical data studies look like and why do we care? Blog post, *Society and Space*. Available at: http://societyandspace.com/material/commentaries/craig-

dalton-and-jim-thatcher-what-does-a-critical-data-studies-look-like-and-why-do-we-care-seven-points-for-a-critical-approach-to-big-data/ (20 May 2014, accessed 2 June 2014).

de Montjoye YA, Hidalgo CA, Verleysen M, et al. (2013) Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports* 3: 1376.

Donovan KP and Martin AKs (2014) The rise of African SIM registration: The emerging dynamics of regulatory change. *First Monday* 19: 2–3.

Eagle N, de Montjoye YA and Bettencourt LM (2009) Community computing: Comparisons between rural and urban societies using mobile phone data. In: *International Conference on Computational Science and Engineering, 2009. CSE'09.* Vol. 4, pp.144–150. IEEE.

Eagle N and Pentland A (2006) Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing* 10(4): 255–268.

Escobar A (2011) *Encountering Development: The Making and Unmaking of the Third World.* Princeton, N.J.: Princeton University Press.

Floridi L (2013) *The Ethics of Information.* Oxford: OUP.

Foucault M (1977) *Discipline and Punish: The Birth of the Prison.* New York: Vintage Books.

Gabrys J (2014) Programming environments: Environmentality and citizen sensing in the smart city. *Environment and Planning D: Society and Space* 32: 30–48.

Gasson MN, Kosta E, Royer D, et al. (2011) Normality mining: Privacy implications of behavioral profiles drawn from GPS enabled mobile phones. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 41(2): 251–261.

Global Pulse 2013. Information sheet. Available at: http://www.un.org/millenniumgoals/pdf/GP%20Backgrounder-General2013_Sept2013.pdf (accessed 3 June 2014).

Gonzalez MC, Hidalgo CA and Barabasi AL (2008) Understanding individual human mobility patterns. *Nature* 4537(196): 779–782.

Graham M (2010) Neogeography and the palimpsests of place: Web 2.0 and the construction of a virtual earth. *Tijdschrift voor Economische en Sociale Geografie* 101(4): 422–436.

Greenleaf G (2013) Sheherezade and the 101 data privacy laws: Origins, significance and global trajectories. *Journal of Law, Information and Science* 23(1): 4–49.

GSMA (2011) Mobile privacy principles. Available at: http://www.gsma.com/publicpolicy/wp-content/uploads/2012/03/gsmaprivacyprinciples2012.pdf (accessed 16 June 2014).

Harley JB (1989) Deconstructing the map. *Cartographica: The International Journal for Geographic Information and Geovisualization* 26(2): 1–20.

Hildebrandt M (2013) Slaves to big data. Or are we? In: Bus J, Crompton M, Hildebrandt M, et al. (eds) *IDP Revista De Internet, Derecho Y Política*, Vol 17, pp. 7–26.

Hosein G and Nyst C (2013) *Aiding Surveillance.* Privacy International. Available at: https://www.privacyinternational.org/sites/default/files/Aiding%20Surveillance.pdf (accessed 20 June 2014).

ITU (2013) *The World in 2013.* International Telecommunications Union. Available at: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013.pdf (accessed 15 July 2014).

Jerven M (2013) *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It.* Ithaca: Cornell University Press.

Kingsbury P and Jones JP (2009) Walter Benjamin's Dionysian adventures on Google earth. *Geoforum* 40(4): 502–513.

Kirkpatrick R (2013) *Presentation at Internet Governance Forum 2013*, 22–25 October, Bali, Indonesia.

Kirsch S (2014) Insular territories: US colonial science, geopolitics, and the (re)mapping of the Philippines. *The Geographical Journal.* DOI: 10.1111/geoj.12072.

Kniveton D, Smith C and Wood S (2011) Agent-based model simulations of future changes in migration flows for Burkina Faso. *Global Environmental Change* 21: S34–S40.

Lucas RE Jr (1988) On the mechanics of economic development. *Journal of Monetary Economics* 22(1): 3–42.

Lyon D (2008) Surveillance society. Presented at Festival del Diritto, Piacenza, Italia, 28 September 2008. Available at: http://www.festivaldeldiritto.it/2008/pdf/interventi/david_lyon.pdf (accessed 14 March 2014).

MacKinnon R (2012) *Consent of the networked: The worldwide struggle for internet freedom*. New York: Basic Books.

Makaryan S (2012) Estimation of international migration in Post-Soviet Republics. *International Migration* 53(5): 26–46.

Manoharan S (2009) On GPS tracking of mobile devices. In: *Networking and Services*. ICNS'09. Fifth International Conference on, pp. 415–418. IEEE.

Manyika J, Chui M, Brown B, et al. (2011) *Big Data: The Next Frontier for Innovation, Competition and Productivity*. Washington, DC: McKinsey Global Institute.

Massey D (1993) Power-geometry and a progressive sense of place. In: Bird J, Curtis B, Putnam T and Tickner L (eds) *Mapping the Futures: Local Cultures, Global Change*, Vol. 1. London: Routledge, pp. 59–69.

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

Michael K and Clarke R (2013) Location and tracking of mobile devices: Überveillance stalks the streets. *Computer Law and Security Review* 29(3): 216–228.

Michael K, McNamee A and Michael M (2006) The emerging ethics of humancentric GPS tracking and monitoring. In: *Proceedings of the international conference on mobile business*, June 2006, pp.34–43.

Morozov E (2012) *The Net Delusion: The Dark Side of Internet Freedom*. New York: PublicAffairs.

Netmob (2013) In: *Proceedings of the 3rd Conference on the Analysis of Mobile Phone Datasets*, 1–3 May 2013. Boston: MIT.

Orange (2015) DEEP – Data for development external ethic panel. Available at: http://www.d4d.orange.com/en/presentation/endowment-and-panel/DEEP-Data-for-development-External-Ethic-Panel (accessed 18 May 2015).

Parr H and Fyfe N (2012) Missing geographies. *Progress in Human Geography* 37(5): 615–638.

Pell S and Soghoian C (2012) Can you see me now?: Toward reasonable standards for law enforcement access to location data that Congress could enact. *Berkeley Technology Law Journal* 27: 117–126.

Pentland A (2011) Society's nervous system: Building effective government, energy, and public health systems. *Pervasive and Mobile Computing* 7(6): 643–659.

Pickles J (1995) Toward an economy of electronic representation and the virtual sign. In: Pickles J (ed.) *Ground Truth: The Social Implications of Geographic Information Systems*. London: Guildford Press.

Pindolia DK, Garcia AJ, Wesolowski A, et al. (2012) Human movement data for malaria control and elimination strategic planning. *Malaria Journal* 11: 205.

Sachs J (2005) *The End of Poverty: How We Can Make It Happen in Our Lifetime*. London: Penguin.

Schuurman NC (2000) Critical GIS: Theorizing an emerging science. Doctoral dissertation, University of British Columbia.

Scott JC (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.

Sen A (1999) *Development as Freedom*. New York: Knopf.

Sharad K and Danezis G (2013) De-anonymizing D4D Datasets. In: *The 13th privacy enhancing technologies symposium*, 10–12 July 2013, Bloomington, Indiana, USA. Available at: https://www.cl.cam.ac.uk/~ks640/files/papers/De-anonymizing_D4D_Datasets.pdf (accessed 5 June 2014).

Sieber R (2006) Public participation geographic information systems: A literature review and framework. *Annals of the Association of American Geographers* 96(3): 491–507.

Solove D (2013) Privacy self-management and the consent paradox. *Harvard Law Review* 126: 1880–1903.

Soto V, Frias-Martinez V, Virseda J, et al. (2011) Prediction of socioeconomic levels using cell phone. In: User modeling, adaption and personalization records. *Lecture Notes in Computer Science* 6787: 377–388.

Stephens M (2013) Gender and the GeoWeb: Divisions in the production of user-generated cartographic information. *GeoJournal* 78(6): 981–996.

Taylor L and Broeders D (2015) In the name of development: Power, profit and the datafication of the global South. *Geoforum* 64: 229–237.

Taylor L and Schroeder R (2014) Is bigger better? The emergence of big data as a tool for international development policy. *GeoJournal* 80(4): 503–518.

Thatcher J (2014) Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication* 8: 1765–1783.

Warren S and Brandeis L (1890) The right to privacy. *Harvard Law Review* 4: 193.

Winner L (1977) *Autonomous Technology: Technics Out of Control as a Theme in Human Thought*. Cambridge, MA: MIT Press.

Zook MA and Graham M (2007) The creative reconstruction of the Internet: Google and the privatization of cyberspace and DigiPlace. *Geoforum* 38(6): 1322–1343.

Linnet Taylor is a Marie Curie postdoctoral fellow in the Governance and Inclusive Development Group, International Development, University of Amsterdam.