In *Freedom and Property of Information: The Philosophy of Law Meets the Philosophy of Technology*. (Mireille Hildebrandt & Bibi van den Berg, eds., forthcoming 2016).

Data subjects or data citizens?

Addressing the global regulatory challenge of big data

Linnet Taylor

> When we share personal data with others, we usually have an expectation about the purposes for which the data will be used.
>
> Article 29 Working Party, 2013[1]

> Even if you are looking at purely anonymized data on the use of mobile phones, carriers could predict your age to within in some cases plus or minus one year with over 70 percent accuracy. They can predict your gender with between 70 and 80 percent accuracy. One carrier in Indonesia told us they can tell what your religion is by how you use your phone. You can see the population moving around.
>
> Robert Kirkpatrick UN Global Pulse, 2012[2]

Introduction

There is an inevitable ethnocentrism currently at play in debates about the power of data and the power over data. 'Global' data problems and solutions are conceived as beginning and ending in regions with meaningful and enforceable data protection regulation, namely the US, the EU and a small group of other OECD countries tightly bound to these regions by trade, such as Canada. Thus the 'we' of the Article 29 Working Party's statement quoted above is clear and justifiable in those regional contexts, but becomes hugely problematic when it is applied to the majority of people in the world. For these 'other' six billion, who live and use technology outside high-income countries (HICs) with clear data protection provisions, even an effort toward purpose limitation such as that quoted above raises some serious questions. Julie Cohen's exploration of the ways in which code mediates between truth and power (this volume) is not only pertinent in the US and EU

where regulation battles are primarily being fought, but becomes even more applicable in places where information flows have even higher stakes. As a Zimbabwean participant in the 2005 World Summit on the Information Society put it when discussing internet freedom, 'if we have no freedom of speech, we can't talk about who is stealing our food' (Mackinnon 2012).

Taking the implications of Cohen's analysis into the context of lower-income countries demonstrates how flows of data mediate power in a myriad ways, even where most people are not yet internet users. Let us take as an example a data analysis competition involving data from Côte d'Ivoire. In 2012, mobile company Orange released a year's anonymised call records from all its Ivoirian subscribers (Blondel et al. 2012). This was the first major release to researchers of this type of 'big data' stemming from a low- or middle-income country, and the first to be labelled a development project. As such the release 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. Since the data were anonymised and blurred, subscribers were not asked for their consent to the release of their data. Yet this was not European data: it was gathered in Côte d'Ivoire, at the end of a year of civil war in that country, and despite its anonymisation researchers were still able to derive communication networks and mobility patterns which in turn identified potentially sensitive ethnic and spatial characteristics and ties. Nor were national authorities invited to outline development aims which might be relevant to this 'development research'. Only one of the 250 research teams who received the data visited Côte d'Ivoire, and the project was governed by no national or international regulations or ethical framework with regard to the privacy of the individuals involved, or the subsequent use or sharing of the data – because such regulations and frameworks do not exist.

For mobile phone users such as those in the Orange project, living in a Sub-Saharan African country with no enforceable regulations dealing with digital data, and using a network provider whose parent company is based in the EU, the debates currently taking place in the EU and US about data regulation may seem irrelevant at best. If a network provider wants to share users' data in their own country or beyond, for research, commercial or 'development' purposes, as is occurring increasingly often, it becomes necessary to revisit the Working Party's statement. Who are 'we'? How do 'we' find out

that 'our' data is being shared, and with whom? Is local knowledge considered relevant in assessing how to prevent harm from such a data release? What happens when the 'others' with whom data is to be shared are located in a political and legal system that is not in dialogue with ours or at least incentivised to find elements of compatibility, as is the case with the EU and US? And, perhaps most importantly, how are technology users to develop an 'expectation about the purposes for which the data will be used' in contexts where they suffer from disproportionate information asymmetries?

This chapter focuses on the question of how big data's potential and risks can be reconciled with regard to the global data landscape. 'Big data' is defined here as any dataset or, especially, combination of datasets which make possible an unprecedented depth and breadth of analysis on a particular question (Schroeder 2014). It is also possible to define big data as more a process than a type of dataset or particular product, where it is characterised by merging, linking and analysing across databases and data types. These characteristics raise new challenges to privacy which are often illustrated more clearly by the context of the 'other six billion', where rules and standards are still in flux and where ethical debates are still being defined. I will argue that considering this context can lead researchers to some new questions with universal relevance for conceptualising privacy, and to some practical approaches to what might be termed 'emerging harms' relating to the misuse of digital data.

It would not be too extreme to say that there are two distinct contexts in which data is emitted, used and shared today: one where regulation exists and has traction, even though issues of jurisdiction and the compatibility of standards may be under strenuous debate, and another where firms and governments are free to gather, process and share digital data under conditions of self-regulation. These contexts may overlap, for instance with regard to intelligence services in HICs, but in general can be taken as a proxy for the divide between HICs and LMICs. I will analyse the implications of these contextual differences with reference to what Scott (1998) has termed *legibility* – the way in which citizens become visible, or legible, to authorities through data collection and analysis. These authorities may be governmental, since one central role of a functioning state is to establish the legibility of its citizens through the collection and processing of statistical data. However, in the era of big data the actors involved are changing as private sector takes a leading role in generating, processing and acting on big data. This shift is giving

rise to another where the state's role in identifying, classifying and intervening merges with that of corporations'[3], a process in which we can distinguish the reproduction of power relations through information technology (or what Deleuze [1992] has termed 'modulation').

This merging of state and corporate interests is progressing particularly freely in low- and middle-income countries (LMICs),[4] where corporate-sponsored 'dataveillance' using big data is being posited as a way to supplement – or replace – underfunded or understaffed state statistical apparatuses (Taylor and Schroeder forthcoming). Here, private-sector dataveillance, and the flow of data between corporations and international institutions such as development donors is facilitated by claims that better data is indispensable if countries are to develop (Jerven 2013). Supporting this argument is a widespread notion that people in developing countries do not care about privacy, or have such different perceptions of it that general standards for privacy-related data protection become meaningless when applied outside high-income countries (HICs).[5] This leads to a situation where increasingly sophisticated surveillance apparatuses are becoming normal in a context of little or no data protection regulation. It also creates the potential for a perfect storm for privacy and data protection rights in LMICs.

The new challenge big data analytics presents to the fundamental right to privacy can be seen most clearly – and in its most ethically complex form – in lower-income countries. This is for two reasons: first, because the mythologisation of the power of big data analytics (Puschmann and Burgess 2014) is leading to the belief among data controllers that access to unprecedented amounts of digital data can solve problems which are in fact rooted in structural and political realities. For instance, claiming that big data analytics can create 'commercially viable solutions to Africa's grand challenges in healthcare, education, water and sanitation, human mobility and agriculture'[6] sounds impressive. Yet these things have not been achieved anywhere else in the world without building stable governance and rule of law, a strong civil society and creating processes of resource management that address the needs of the majority rather than the elite. It is unlikely that such problems can be solved by giving unregulated organisations license to collect and share data as freely as possible. Furthermore this discourse arguably gives rise to a false dichotomy where the right to privacy must be weighed against such emotive concepts as fighting poverty and disease. This ethical minefield is situated in a regulatory

context where states have yet to develop comprehensive standards for data protection that relate to today's digital data, along with the capacity to identify data misuse and to enforce those standards against international corporations. If, as Cohen writes (2012), privacy 'is an indispensable structural feature of liberal democratic political systems', should this be allowed to imply (as many currently believe) that it is a less fundamental right in places with different structural characteristics or more urgent basic needs?

This chapter begins from Cohen's challenge (this volume) to examine the types of power emerging in relation to the information age. Going beyond the problem of state power, I reframe the current privacy and data protection gap in LMICs as a lack of effective regulation of corporate and development actors, and ask how privacy theory needs to stretch to accommodate this scenario. The legibilities created by the new data empires are not designed by states, and thus do not aim at representing citizens or supporting legitimate state interests such as the rule of law or taxation. Instead they are aimed at making citizens better data subjects and consumers, or at best, better subjects of development interventions. This gives rise to a modulation of the kind Cohen describes, but one which operates at a societal level rather than targeting particular groups or behaviours. The use of observed and inferred data (Hildebrandt 2013) in countries where official data collection has always been limited gives rise, I argue, to a condition of ever-increasing legibility without better political representation. This implies that the common argument offered with regard to privacy in developing countries – that it is a luxury the poor cannot afford – may only be a convenient way to reproduce existing power relations. It gives corporate and development entities the opportunity to bypass human development (including political expression, security, health and education) in favour of forms of economic expansion which chiefly benefit a core of richer countries rather than the local subjects of development.

One way to address this problem, I will argue, is through a regulatory approach that assumes that privacy is a fundamental right even in situations where it is inconvenient to economic development priorities. In this context, I argue that data protection regulation is an important instrument for protecting privacy in its various dimensions, and that holding HIC-based corporations to account for their actions worldwide could begin a global dialogue about privacy focusing on more than just high-income countries. I posit that political and economic context are essential to

understanding how to establish the right to privacy, since they define the tensions between corporations, states and regulators. The contribution of this chapter is to propose both a rationale and an analytical framework for conceptualising and enacting a global right to privacy in the era of big data, and to do so without engaging in a cost-benefit analysis which privileges transactional interpretations of privacy (such as privacy self-management, the idea that the individual can manage and control flows of data about themselves and that they will trade certain aspects of privacy for other advantages) over fundamental rights. To do this, I frame big data approaches as a socio-technical system which will have different features in different locations but which can be approached with a unified ethical framework and which can be subject to regulation anywhere in the world. I will use examples to illustrate the heightened stakes with regard to privacy and data protection in LMICs, where harms to individuals may potentially be much greater than in HICs, and the ways in which data regulation may take this imbalance into account to provide a system which can grow with the expansion of technology instead of needing to be reinvented for every new location.

EVOLVING PRIVACY DEBATES WITH REGARD TO LOW- AND MIDDLE-INCOME COUNTRIES

Cohen (this volume) has pointed out that traditional legal institutions are increasingly being sidelined in terms of regulating and adjudicating data and information-related governance issues. Instead, she has noted, new networked and often private-sector systems of governance are evolving around information processes. Looking at the international instruments which might be expected to set the initial parameters for data governance in LMICs, this marginalisation is evident. Although privacy with regard to personal data has been established as a fundamental right in international declarations, it has yet to be set out in national law or enforced in practice across large portions of the world. As a right privacy is clearly defined in Article 12 of the Universal Declaration of Human Rights (1948) and in article 17 of the International Covenant on Civil and Political Rights (1966), but regional charters have taken longer to establish it and national laws even longer. For example, the Organisation of African Unity Charter on Human and Peoples' Rights (the Banjul charter, passed in 1981) makes no reference to privacy, nor

does China's constitution (White and Case 2012). Although more recent articulations of rights are starting to take privacy into account, such as Article 30 of Hong Kong's Basic Law (1997) and Article 14 of its Bill of Rights (1991), and article 7 of the East African Community's Draft Bill of Rights (2009), the most recent analysis of global data privacy laws (Greenleaf 2013:11) shows that the overwhelming majority of states without such laws are low- or lower-middle-income countries. For example, in Sub-Saharan Africa only 8 states out of 55 have data protection laws (ibid). Furthermore, there is little agreement over how data which does not identify individuals but which nevertheless enables proxy forms of identification such as by location and activity should be regulated (Irion and Luchetta 2013), and even in the field of traditional 'personal data' there is disagreement about which data types are most sensitive and why, and therefore about how to protect them.

Even as data protection legislation spreads, practical problems unfold as to how privacy principles apply practically (in terms of rules, regulation and standards rather than international norms or national legislation) to people in lower-income countries in the era of big data. First, privacy has so far been conceptualised in ways which are most applicable under conditions of strong statehood and rule of law, mainly as protecting individuals from *too much* statehood and law, i.e. surveillance, persecution by the state, and oppressive laws (e.g. Solove 2006, Nissenbaum 2009), or, more broadly, protecting a 'socially constructed self' from interference with their self-determination (Cohen 2013), with the assumption that the individual is positioned as the most important unit in society. Although these are desirable preconditions for legislating privacy, they may not be realistic or broad enough to encompass how privacy may be conceptualised in LMICs. For instance, the Banjul Charter frames some rights from the perspective of the individual and some from that of the group, examples of the latter being 'protection of the family and vulnerable groups', the 'right to free disposal of wealth and natural resources', and the 'right to economic, social and cultural development' (Banjul Charter 1981).

There is thus an interplay between conceptualisations of privacy and the reality of governance, both of data and more broadly. A related problem is that the distribution of power over digital data is evolving differently in lower-income countries, where the dynamics of economic development are making corporations (and in some places international organisations) more important than national governments. The landscape of

data protection is therefore weighted toward remote, multinational actors rather than domestically controlled entities. States are further disadvantaged by the technical resources and capacity necessary to manage and regulate big data, and therefore tend to contract in digital communications expertise (oAfrica.com 2013). Thus corporations become the main actors providing connectivity infrastructure, gathering and managing data, rather than governments.

Although data protection standards in LMICs tend to be low, digital technology use is rising exponentially and providing a rich market for data with few controls. Everywhere in LMICs multinational technology corporations can be found stepping in to play the kinds of roles that in HICs were initially played by national monopoly providers who were easier to regulate. Telephony in Sub-Saharan Africa is a prime example: due to the lack of resources for states to invest in landline infrastructure, once mobile phones became available they took hold rapidly. The proportion of people in the region with a mobile subscription rose from 12.4% to 63.5% across the region between 2005 and 2012 (ITU 2013). This dynamic also increasingly governs financial activity: day-to-day financial transactions in LMICs such as microfinance payments and individual transfers are increasingly taking place over mobile phones through e-payment and instant money systems (REF). The system with the most widely-publicised and exponential growth is Kenya's MPesa money transfer network, where usership has risen from 19,600 in 2007 to more than 15 million clients in 2014 (Safaricom 2014, Forbes 2014). Just as mobile phones have leapfrogged landlines and mobile transfers traditional banking systems, digitally based biometric identification systems are similarly gaining traction as an alternative to state-issued identification documents: in India, for instance, the biometric Unique Identification scheme was set up by corporate experts under circumstances described by Greenleaf (2010) as a 'privacy vacuum', and with only a fraction of the population registered is already the world's largest biometric database with more than 415 million records (Srinivasan and Johri 2013).

Each of these examples of corporate leadership helps illustrate both how fast technology adoption may move in lower-income countries, and how little opportunity there is for conceptualisations of privacy and data protection to keep up with it. In contrast to Sub-Saharan Africa's lack of privacy legislation, 48 out of 55 countries have adopted compulsory SIM card registration (Donovan and Martin 2014), often with

connections to central identification databases, so that any activated SIM card is now linked to an individual through both corporate and governmental information systems. Smartphones, which are more likely to identify their user through apps 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% (ibid.), but has risen six-fold since figures became available in 2010.

These new networks of information and identification have already seen significant privacy risks and some breaches. The release of mobile phone call records by mobile provider Orange, described in section 1, although anonymised to established HIC standards, was noted by privacy researchers to offer broad opportunities to identify the movements, communication networks and activities of Ivoirians during a fragile postwar context using standard methods which could easily be used to inappropriate ends (Sharad and Danezis 2012). As these researchers point out, the risks to data subjects in such a case of big data analysis may accrue on a more general, group level – such as through the identification of a village rather than an individual. Kenya's MPesa payments system has seen an actual breach of client data protection (TechMtaa 2012), where would-be political parties have 'bought' their way into the financial transactions database, harvesting the names of users both to register them as party members and thus gain official standing as parties, and to send political messages during elections (a particularly dangerous abuse of personal data in a country with a history of extreme election-related violence). Meanwhile, in India's biometric ID system it has recently become public that multiple foreign contractors have been involved in building and running the database, and that these include firms from the US who are obliged under the Patriot Act to make their data (in this case Indians' private information) available to US intelligence services (New India Express 2014). These examples highlight how corporate data practices and alliances are likely to be the main sources of privacy violations regarding digital data in LMICs, and that corporations must be a central focus of regulation and reasoning with regard to digital data. I frame them as privacy risks, although they also inevitably imply broader issues of data protection, because they illustrate Warren and Brandeis' early articulation of the right to privacy (1890) as 'the right to be let alone'. In a context of data analysis for policymaking and intervention (as in the Côte d'Ivoire case), or for sending unsolicited political messages in a climate of unrest and violence (as in Kenya), the right to be let alone becomes the overarching issue.

Another risk to privacy is the involvement of international development actors – Non-Governmental Organisations (NGOs), the United Nations, and bilateral development partners – all of which desire more and better data to inform and evaluate development interventions. One high-profile example is the UN's Global Pulse initiative, whose mission is to encourage 'data philanthropy': the donation of corporate datasets by large firms operating in LMICs to inform development actors. Through big data analytical methods Global Pulse aims to predict economic shocks and provide information to humanitarian responders in LMICs – but it has also voiced a need for new ways to think about privacy in this context (Robert Kirkpatrick, UN Global Pulse,[7] 2013):

> We think big data here is the greatest opportunity to present itself to global development in many, many years. Unless you fail to protect privacy in the process, in which case this may be the greatest threat to human rights the world has ever known.

There is reason to worry: the work of Sharad and Danezis (2012) along with that of other researchers working on LMIC privacy issues (e.g. Taylor 2014) points out how various bedrock concepts of data protection may need to be reconceptualised as data flows become multi-layered and multidirectional. As Solove (2013) has pointed out, in the era of big data analytics, the 'identifiability of data depends upon context.' Anonymisation may be an even less reliable approach to data protection in LMICs than it is in HICs, and purpose limitation is already becoming a flimsy concept as 'development' and 'the public good' are used to justify uses of data for research and prediction far beyond the knowledge of ordinary technology users. Furthermore, these bedrock concepts of data protection, along with the EU's standard that data processing must have a legal basis (EU Data Protection Directive 1995) are rendered effectively meaningless by the near-impossibility of enforcing any legislation based on them where the data originates in LMICs. Although there have been attempts to protect HIC citizens' data when it is processed in third countries in the form of safe harbour provisions and adequacy agreements, conversely no provisions have been developed to guard the integrity of LMIC citizens' data when it is analysed by HIC organisations under the rubric of development or humanitarian action, as in the Côte d'Ivoire example above.

The challenges noted above add up to a scenario of serious power and information asymmetries affecting citizens of LMICs with regard to their digital data. To summarise the main sources of these imbalances:

1. *Data maximisation*. Big data is seen as a way to engineer development to overcome otherwise intractable problems,[8] and as a solution for structural and political problems[9]. The public commentary so far on digital data in LMICs focuses mainly on access to data (Green 2014), open data (World Wide Web Foundation 2014) and development data (Global Pulse 2013) but not on restricting or awarding rights over data. This leads to widespread acceptance of data maximisation (in contrast to the standard privacy principle of data minimisation [Schermer 2011]), in the interest of more effective data mining. In turn this normalisation of seeking as much data as possible leads to a false dichotomy being drawn between 'privacy' and the interests of the poor (with neither clearly defined), and to the almost religious idea that if enough data is made available, 'solutions' to complex problems such as poverty and disadvantage will be found. Data maximisation, however, also presents a greater risk to privacy over the long term as the big-data approach of aggregating different data sources raises the likelihood that apparently innocuous information about people will become sensitive in connection with other datasets (Solove 2013)

2. *Differing parameters for conceptualising privacy.* As noted in the examples cited above, and particularly the Côte d'Ivoire case, the parameters for defining privacy may shift with regard to LMICs. For example it may become more important to consider group welfare and harms alongside individual identifiability, something that challenges the current focus of privacy scholarship and legislation on Personal Information – and possibly also on the conflation of the concepts of privacy and data protection. For instance, being identified as part of a community that is prone to HIV infection or that is likely to be forced to migrate across international borders is potentially valuable in that it allows protective intervention, but also holds risks where those communities may be subjected to unwarranted or preemptive action by outsiders based on predictive analytics. Equally, the risks involved may start with individuals' right to be let alone, but also incorporate problems best articulated by the broader concepts of data protection that also regard the presumption of innocence. However, the main current articulation of data protection principles (the EU 1995 Data Protection Directive) encapsulates exactly this

conflict. The DPD states that data should not be used to profile where this can cause discriminatory action (article 8.1), but also notes that this does not apply where 'data processing is necessary to protect the vital interests of the data subject or of another person' (8.2[c]). In the context of extreme poverty, disease and the other challenges in LMICs which might prompt data analysis by outside organisations, it is fair to assume that an analyst trying to comply with available data protection rules will believe herself to be acting to protect the vital interests of data subjects, and therefore to be exempt from restrictions.

3. *Distributed governance scenarios*. Information and power asymmetries with regard to digital data are created and reproduced where states lack the resources or capacity to effectively regulate data flows and to enforce rules and standards. Where this occurs there is a corresponding empowerment of corporations to connect people, gather and analyse their data, and a feedback loop where the state is not incentivised to rectify the asymmetries as long as corporate self-regulation appears not to be causing visible harms.[10] This imbalance in governance is supported by the almost universal promotion of the Public-Private Partnership (PPP) as a mode of delivery for development interventions and decentralisation initiatives in LMICs, seen as the best way to keep donors' costs down and increase efficiency in humanitarian and development activities (Miraftab 2004). In a context of strong corporate empowerment and limited state involvement in digital data flows, individuals are also less likely to be informed of their rights with regard to their data, to have experience of or capacity for actively managing their electronic privacy, or to know through the media if their data is being misused.

4. *Higher stakes*. The categories of harm set out in Solove's taxonomy (2006) are a useful starting point for any analysis of the stakes involved in privacy and data protection: he distinguishes between more ad-hoc 'reputational harms' and longer-term 'architectural issues' which are more relevant to today's uses of digital data. However, Solove's examples of possible future harms illustrate perfectly the different considerations that are appropriate for HIC and LMIC contexts. Solove refers to the risk of 'being victimized by identity theft or fraud', and the way that 'People's behavior might be chilled [by surveillance], making them less likely to attend political rallies or criticize popular views.' These risks are real, but in an LMIC context it is easy to see how possible harms may go beyond those conceivable for citizens of HICs. For example, exposing

ethnic, religious and political affiliations or identifying dissidents' social networks has been shown to constitute an actual risk of harm for individuals in authoritarian states (Mackinnon 2012). Similarly, researchers have shown that it is a relatively simple process to determine people's sexual orientation through the public information disseminated by social networking websites (Jernigan and Mistree 2009), something which is relatively innocuous in the academic context but highly sensitive in the many countries where harsh laws against homosexuality combine with low standards for evidence.

Borgesius (2014) points out that according to Article 8 of the EU 1995 Data Protection Directive, 'data protection law has a stricter regime for "sensitive data", such as data revealing racial or ethnic origin, religious beliefs, and data concerning health or sex life.' The examples offered above suggest that in relation to big-data-type analytical processes, not only is anonymisation insufficient to protect what Floridi (2013) might refer to as people's 'ontologically constitutive' information, but for those living at risk of violence or persecution, apparently anonymous data such as social network structure may reveal highly sensitive and potentially harmful information about them which under the EU's rules should be protected. Again, this blurs the distinction between privacy and data protection more broadly, where the right to be let alone may be the primary issue, but the way in which that right can be claimed may involve issues more associated with data protection, such as non-discrimination and transparency.

The next section outlines how a global approach to data protection might be created by taking into account the factors explored above, suggesting that data protection needs to be treated primarily as a regulatory problem rather than a holistic one of conceptualising, advocating and realising global norms with regard to privacy.

CAN PRIVACY AND DATA PROTECTION PRINCIPLES APPLY BEYOND HICS?

The main debate on privacy and data regulation is taking place within a core of HICs, mainly the EU states, the US and Canada. Many other countries are included in these debates as legal satellites to which the US/EU approaches to data protection are radiating due to those countries' pragmatic adoption of similar codes and standards due to trading and colonial relationships (Greenleaf 2013). A broader view, as set out in the previous section, shows however that there are extensive regions of the world which may be

conceiving of people's rights in relation to data, corporations and the state (if they do so at all) according to very different rules and precepts. Bradford (2012) argues that the EU's effective monopsony power as a consumer bloc combined with its strong regulatory capacity mean that other countries effectively have to adopt EU regulations on data protection and privacy, even if they have their own differing systems in place. However, the uses of data described above (development-related research such as the Orange study, or India's domestic biometric identification system) are effectively outside the realm of the EU's influence because they involve functions that are not directed toward engagement with the EU.

There is thus a substantial category of 'other' comprising a majority of the world's technology users, and increasingly including a variety of hybrid uses of data involving multiple actors only some of whom may be subject to regulation or influence by existing data authorities. Given the governance factors set out above, these territories far from regulators' grasp also represents an increasingly viable place of refuge for firms or organisations which are interested in stretching the boundaries of permissible data use. It is in the context of this broader picture that the debate between US and EU data protection principles and laws should be framed analytically. If digital technologies are being used worldwide and data emitted from every country on earth, it is inevitably the global context that will determine whether data privacy provisions such as the forthcoming EU Directive will play a role in creating just and fair uses of data.

The realist argument offered above is important to consider because it affects the way laws can be enforced with regard to the transnational collection and use of data, practices which are almost universal amongst large technology firms. If the actual global scenario with regard to data protection frameworks (US, EU and 'other') is not taken into account, both international coordination and the ability to understand the evolution of regional regulatory frameworks suffer. Privacy as a fundamental right (the dominant perspective in the EU according to the Charter of Fundamental Rights) versus privacy as a transactional attribute which can be managed and weighed against other benefits (dominant in the US due to its more piecemeal approach to data regulation) are only two of many ethical perspectives which may inform data protection regimes. Others include customary or common law frameworks which may be biased toward communal rather than individual rights (loosely, the African idea of Ubuntu, or collectivity, see for

example DuPlessis 2011); the view that privacy has a chiefly instrumental value as a way of preserving social order (China's evolving articulation of the problem, see Yao-Huai 2005) or that groups have bounded and contingent rights to privacy depending on their socioeconomic status and country of citizenship (as applies to certain categories of migrants, resident foreigners and the poor worldwide – see Gilliom 2001).

The table below compares the central aspects of privacy with regard to data in HICs versus LMICs. The main elements covered are the principles for regulation; the governance conditions which determine the degree of traction any regulation may gain, and the types of risk associated with each regulatory environment.

Privacy in HICs vs LMICs

	High-Income Countries	**Low- and Middle-Income Countries**
Regulation	'information relating to an identified or identifiable natural person' (EU data directive '95); 'personally identifiable information' (McCallister et al. 2010)	Little regulation; status quo weighted in favour of surveillance and control, e.g. 8 of 55 countries in Africa have data protection laws, mainly analogue (Greenleaf 2013), but 48 have SIM registration (Donovan and Martin 2014)
Governance	Centralised governance: rule of law, more accountable states, enforceable limits on corporations	Distributed governance: limited statehood, less accountable states, few enforceable limits on corporations
Analytical challenges	Anonymisation/blurring of data; data aggregation level; access provisions; jurisdiction of data protection rules	HIC concerns PLUS a lack of supplemental country data may lead to inaccuracies in remotely-conducted data analysis
Risks	Individual harms: discrimination, tracking, identity theft, unwanted marketing	HIC risks PLUS displacement; blocked mobility; violence; political repression PLUS group harms

The table helps illustrate, as Greenleaf points out, that 'a law on the books is not to be confused with effective privacy protection' (Greenleaf 2013). Notwithstanding the presence in a country of a comprehensive data protection act, or privacy legislation more broadly, it may not be realistic to lean on state legislation and enforcement. For example, for a citizen of Côte d'Ivoire whose data was involved in the Orange Data for Development challenge of 2013, the ability to contest the reuse and sharing of their data for broad research purposes, or the way in which it has been anonymised, involved first gaining the knowledge that the data had been used; second, understanding how this use might impact them; third, having a local data protection authority to complain to; and fourth, for that authority to have the resources and influence to hold data processors accountable in an EU or US court. These factors are further problematised by the fact that the dataset was, according to European, US and industry standards (e.g. EU Directive 95/46/EC [1]; GSMA 2011), 'properly anonymised', which would have rendered data protection law inapplicable if it applied to the Ivoirians' data – which it did not.

This example suggests that the inclusion of a more global perspective in thinking about data regulation is not only wise, but pragmatically important. The problem of privacy is similar to that of pollution: if effectively regulated with industrialised countries in mind, the problem will simply move elsewhere to where laws become principles and thus unenforceable. This is already occurring with data self-regulation by corporations in the sphere of international development (Taylor and Schroeder 2014), where the public announcement of good intentions allows data sharing and access which would not be permissible under any other conditions. Carly Nyst of Privacy International has said that (Nyst 2013):

> it would be impossible for any European-based enterprise to collect location data on minorities, require SMS-reporting about drug adherence, or establish large databases of sensitive information without safeguards.

Yet this is currently the scenario for data use and reuse in most LMICs, either under the banner of standard corporate data processing practices, as Kirkpatrick's quote at the start of this chapter indicates, or in the course of devising interventions in the name of development.

Nissenbaum (2009) has written that one important function of privacy is to take control away from states. Solove has written that privacy guards self-determination (2006). Despite the evident usefulness of their perspectives, they contrast strongly with the way that big data analytics are evolving with regard to LMICs. In the global landscape there are multiple actors already involved in analysing personal data, ranging from multinational technology giants to development organisations and the UN, for purposes ranging from economic empowerment to humanitarian and development interventions. This highly varied landscape is more like that described by Neil Richards when he notes that 'It might seem curious to think of information gathering by private entities as "surveillance"… [yet in the era of big data] government and nongovernment surveillance support each other in a complex manner that is often impossible to disentangle'. A similar problem is outlined by Cohen (this volume) where she notes that 'Debates about state censorship … represent only one piece of a larger puzzle, which concerns the extent to which global circuits of information flow are settling into patterns that serve larger constellations of economic and political power.'

The range of actors involved in the power dynamics of dataveillance in LMICs is particularly broad, ranging from states interested in surveillance to NGOs working exclusively with the aim on public benefit. Any meaningful data protection established in LMICs therefore has to take this range of actors into account, and allow for some uses of data which are potentially invasive of privacy, particularly in the humanitarian sphere, while also strongly curtailing practices of data maximisation and surveillance leading to social control and the suppression of dissent. Moreover, there is no guarantee that these two ends of the spectrum will not sometimes merge. Unregulated actors with privileged access to personal data, even those with infinite goodwill, will not always use it in ethically desirable ways or ensure that it never reaches those who may have worse motives: the development and humanitarian communities certainly have no history of being right all the time. Yet data privacy is entirely a matter of self-regulation among these groups, nor have ethical frameworks yet been developed to guide data use in these cases.

The following section looks at principles and practical ways to address this scenario of self-regulation, with particular attention to the tension between privacy self-management and workable solutions for the LMIC contexts analysed here.

CREATING INTERNATIONAL CHECKS AND BALANCES FOR DATA FLOWS

The examples outlined here have illustrated the particular difficulties facing citizens of LMICs in resisting unwarranted uses of their digital data. First, the political and economic architectures within which big data is produced and analysed in lower-income countries – particularly the extra-national and transnational flows of data characteristic of multinational corporations and development organisations – tend to place citizens at even more of a disadvantage than people in HICs in terms of controlling what happens to the data they emit. Second, if an unwarranted use of data is identified, a lack of local legislation or rules may prevent citizens from seeking recourse. Such action has to be backed up by local authorities in the case of the transnational use of data – something which is both difficult where lower-income or smaller states are structurally disadvantaged in terms of international influence, but is also a challenge for those equipped with the full force of the law and acting from within the EU (as can be seen from Max Schrems' attempt to use EU privacy regulations to sue Facebook in the EU [Press Association 2014]).

These structural problems are what Solove (2013) refers to when he argues that privacy self-management is not an option for ordinary technology users. They also support Cohen's statement (this volume) that the new pathways of data governance are frequently hidden within the auspices of the private sector, and that they are not even theoretically accessible to citizens in the way that national laws and regulations should be. The conditions of corporate self-regulation and little state intervention or power over data flows which currently prevail in most LMICs strongly support this argument. If US citizens (as argued by Solove, ibid.) and EU citizens (as argued by Borgesius 2014) cannot be realistically expected to manage their own privacy with regard to digital data, what should be the solution in places with greater structural disadvantages and information asymmetries? A global perspective makes it clear that privacy self-management is an approach that inevitably favours certain economic and geographic groups over others, namely those who are continuously connected to the internet by smart devices and are technologically well-informed, and are therefore more able to regulate their own data effectively or to seek legal redress in case of misuse. A large proportion of

the 'other' 6 billion people currently emitting digital data, then, are not only unprotected but, under current assumptions, unprotectable.

Even where people can be effectively informed of their rights with regard to their data, a further problem arises: various solutions which have been proposed that weight the informed consent process in favour of the individual seem a bad fit for the context of development/humanitarian data uses described here. For instance the idea of privacy self-management via a 'personal data store' (Hardjono et al. 2014) or creating greater user awareness through 'visceral notification' tactics (Calo 2012) cannot be considered appropriate in a context where people have too-simple devices or intermittent connectivity. There is also a problem where the data is to be used for health or humanitarian purposes, where data processors will almost inevitably classify their analysis as important to the wellbeing of the data subject and therefore justifiable even without ethical checks or user permission.

Moreover, the framing of 'big data' as a solution for problems which have a structural and political origin makes it more possible that regulation will be de-prioritised in favour of a race to practice the most innovative methods of data analysis (as with Orange's D4D project, described above). This framing also distracts from the other facet of big data in LMICs, and everywhere else – as a form of economic activity relating strongly to growth and innovation. This facet, however, is also liable to misuse in LMICs if regulation is not established and made enforceable. There has been a tendency among development donors to attempt to replicate economic models which have worked in HICs as a way to alleviate problems of governance and production in LMICs. Perhaps the clearest example of this was the imposition of Structural Adjustment Programs (SAPs) in the 1980s and 90s, which attempted to impose a neoliberal market model on low-income countries, and which resulted in the forced marketisation of public services, currency devaluations and, in consequence, political instability and elevated mortality rates amongst vulnerable groups. Programs such as these, introduced without considering local and contextual factors or the limitations of local systems and resources, illustrate the possible long-term consequences of treating development as an engineering problem and data as the essential input.[11]

One classic example of data-driven sorting and categorising with profoundly negative consequences was the passbook system which formed the bureaucratic underpinning for apartheid policy in South Africa (Kahn 1966: 91):

> Every African over sixteen must have on his person what is called a reference book, a bulky document measuring five by three and a half inches and containing ninety-five pages. As a rule, it is only Africans who are stopped by the police and asked to produce their passes. 'The African must be a collector of documents from the day of his birth to the day of his death', says a publication issued by the Black Sash. His passbook must contain particulars about every job he has had, every tax he has paid, and every X-ray he has taken. He would be well advised, the Black Sash has suggested, not to let himself get too far away from his birth certificate, baptismal certificate, school certificates, employment references, housing permits, hospital and clinic cards, prison discharge papers, rent receipts, and, the organization has added sarcastically, death and burial certificates.

In cases where people's data was mishandled or misreported, Horrell (1960) notes, the individual became liable for any mistakes in the dataset and was subject to a prison term for misidentifying themselves, even in cases where they were illiterate and had no idea what their passbook contained. A better example of information asymmetry is hard to find. In the context of development policy-making, people's digital traces may become the contemporary equivalent of the pass book – a complete record of movements, activities and social connections that may be subject to search by a myriad authorities in the name of care and protection.

Structural factors such as poverty and lack of political rights serve as confounders in an analysis that considers how privacy can be a fundamental right everywhere. Yet they should be seen as complications, not contra-indications. Activism occurs in relation to connected types of fundamental rights even in situations where they are strongly contested. For example, the right to self-determination is not lessened because certain groups believe women should be subject to men in many societies; similarly the right to equality before the law can coexist in places such as India with caste systems which restrict the ability of individuals to claim that right. Nevertheless, structural complications are real and meaningful, since in the case of privacy they have contributed to a lack of necessary civil society pressure towards digital privacy in most LMICs. The example of

India, however, suggests that the evolution of public pressure in favour of digital privacy will evolve where a high-profile use of personal data, such as that country's biometric ID project, combines with highly visible and publicised risks to privacy and actual harms.[12]

Given that civil society will respond at a slower pace than the uses of digital data can evolve, and given that national legislation (where it exists) will only gain traction where structural factors support it, both nationally and internationally, what kinds of solutions are useful in the current landscape of big data analytics? Solove (2013), writing of the US and EU context, advocates keeping privacy self-management in play, both because it plays an important role in initially making people aware of their rights with regard to their own data at the point where they begin using a new technology product or service, and because in some cases it is sufficient to the privacy management task at hand. He suggests moving toward what he terms 'paternalistic' approaches, which are relevant to all data use contexts: structuring people's choices so that they lean toward choosing privacy over openness with the data they emit; limiting the role of self-management to what is clear and understandable rather than expecting people to micro-manage; moving privacy decisions 'downstream', i.e. coming back to people for consent when new uses of their data are going to occur; and establishing default rules for data processors according to 'basic privacy norms'. These are all useful, but are inevitably partial in the contexts described above, where consent decisions are easily rendered more complicated by the belief that certain uses of data may contribute to overall wellbeing, or by lack of access on the part of data processors to subscribers in remote areas or those who spend most of their time offline.

For more global solutions, it becomes necessary to move toward a more layered approach to data management, where checks and balances are established and applied at the point of the technology or service's origin, and again at the international level in the case of transnational data sharing and reuse. For example, some possible approaches to take into account the global nature of data flows might include:

1. *Trade restrictions*. Imposing export controls on data processing technologies developed in the EU or US, as occurs with dual-use technologies with possible military ends under the 1994 Wassenaar Arrangement. This has already begun to take place with various surveillance technologies being placed on the Arrangement's 'control list' (Privacy International 2013). It is possible to posit that as well as the data-gathering surveillance

technologies that have been the focus of Privacy International's advocacy, particular data analytic technologies may also be eligible for such treatment.

2. *Redefining checks and balances*. Moving from fixed definitions of personal information and potential harms to a more flexible, easily contextualised, risk-based approach to privacy. This would involve corporations having to submit a form of Privacy Impact Assessment (PIA) to local data protection authorities in their headquarter countries for their work in LMICs, as if they were going to use personal data belonging to US or EU citizens, but in an adapted form. These reworked PIAs would include analysis of local factors in the data's country of origin, preferably involving natives of that country with local knowledge, and would provide information to weigh the immediate benefits of a particular use of data against the potential future risks of re-identification and harm. This approach, though contested (Cavoukian et al. 2014) would provide an initial impulse for corporations to consider the potential impacts of their use of data outside the established regulatory constraints, and is currently the subject of discussions led by the Center for Information Policy Leadership and the World Economic Forum, among others.

3. *Creating institutional reference points*. Such a risk-based approach also suggests real-time, case-by-case institutional support for decisionmaking about data uses that is enforceable against corporations in their place of origin if they misuse data. Several models are available, and though all are far from perfect, a combination would provide greater protection to people outside the HICs than is currently available.

> i) Institutional review boards. These have been used by humanitarian projects looking to perform big-data analytics in emergency situations (for example, the epidemiological analysis of NGO Flowminder during the Haiti cholera epidemic beginning in 2010, which used detailed mobile phone call data from Haitian providers, was reviewed by a Swedish ethics committee at the main researcher's home university)
>
> ii) International ethics committees. These sector- or field-specific committees can be found in areas ranging from medicine (the WHO's Research Ethics Review Committee is one example) and livestock research (the Institutional Research Ethics Committee of the International Livestock Research Institute) to accountancy (the International Ethics Standards Board for Accountants). These

ethical bodies gain traction for their decisions from a combination of factors: strong network incentives to be part of professional bodies and associations which can then become subject to the ethics committee's governance, and the ability to restrict funding or de-certify members who do not play according to the rules established. Both these conditions are problematic with regard to multinational technology firms, however. It has been demonstrated (most notably by the lack of growth in the Global Network Initiative's membership) that such firms are more likely to join associations when forced to by political pressure, while they have the financial and political autonomy to withdraw at will from any inconvenient international arrangements.

iii) Add-ons to the functions of existing international organisations. Although many international organisations with links to LMICs are already involved in the kind of big data analytics described above (including the UN and OECD), there are many organisations with both technical capacity and extensive international networks which could take on a standard-setting or ethical role with regard to data processing in LMICs. Election monitoring provides a possible model: here, the organisations involved are primarily regional bodies (the Organization of American States, the African Union, the European Union and the Council of Europe) but there is also involvement from other types of grouping (the Organization for Security and Co-operation in Europe and the Commonwealth Secretariat) which could provide a model for the registration and monitoring of large-scale data processing projects, and call offenders to account in the international public sphere.

These more protective approaches may be termed paternalistic in the sense of Solove (2013) because they aim to limit harm in specific ways, in advance of data access or processing. They become appropriate, however, in cases where information asymmetries are not rectifiable by other means due to the kinds of structural constraints outlined above, involving both states and individuals. These approaches might address both of the problems of data governance set out by Cohen (this volume) by integrating control over information flows in traditional legal and rights systems, and also by helping those systems evolve to deal with the challenges posed by new types of data generation and processing. Avoiding such protective strategies in the hope that individuals will become

ready to self-manage their privacy through advances in technology and connectivity is unrealistic both because such an evolution is unlikely to occur in the short term, and because the increasingly complex ways in which data are shared and reused are leading to a scenario where it is impossible for anyone to be fully aware of and manage the data they emit, regardless of their level of technological access and sophistication.

CONCLUDING REMARKS: EXPANDING DATA PROTECTION FROM 'WE' TO 'EVERYONE'

Empirical analysis shows that people in LMICs do not have the same access to data protection as those in HICs, and that this is both actually and potentially harmful in a variety of ways. The 'we' of data protection, ascribed as if it were universal, in fact denotes citizens of the EU countries, the US, and a few other HICs. As importantly, this 'we' also tends to denote the elites who can use the range of options currently available, from privacy-self-management tools to legal recourse if necessary. If 'informational capitalism' (Castells 1996) is a reality, as asserted by Cohen (this volume), then those situated nearest to the global centres of the tech economy are likely to benefit from the economic returns created by the new flows of data. In contrast, citizens of LMICs have been largely excluded from the economic benefits of the new data economy, and, further, have less protection from its negative and exploitative effects. Moving beyond the convenient idea that people in lower-income regions are less capable of conceptualising privacy clearly, of advocating for it, or of enforcing rules should they become available, it is clear that advances in big data availability and analytics are everywhere overtaking people's ability to resist the misuse of data about them – and that under certain political and economic conditions there are structural barriers to resisting this process effectively. In order to be able to assume that the right of privacy is indeed fundamental, i.e. applicable to everyone, everywhere, it is necessary to understand those structural barriers, and how to mitigate them with additional types of protection that will allow people to claim and enforce their rights regardless of location.

This chapter calls attention to the ways that existing systems may evolve to deal with new data problems, and the ways in which that evolution may be fostered in the immediate term. It addresses Cohen's point (this volume) that the emerging systems for

data governance are often based more in the private than the public sector, and that these tend to bypass traditional legal and rights instruments which have not yet evolved to deal with new technologies. The points made here address the practical question of how to create an international architecture for the responsible use of data about individuals. I have outlined the problems that stand in the way of such an architecture: first, that data from developing countries is being shared, reused and analysed in ways that are effectively far beyond the jurisdictional reach of current privacy and data protection regulations; second, that these uses are seen by data processors as justifiable on humanitarian grounds and therefore not subject to existing ethical frameworks; and third, that current assumptions about ethical data use may not fit with the way that rights are being conceptualised in developing countries. I have argued that the way to help such an architecture for the ethical use of data to evolve worldwide is to treat multinational corporations processing data as if they were liable for their actions in a similar way worldwide, and to provide clear and usable ways for them to do this.

The analysis above has highlighted how structural barriers to the establishment of a right to privacy and data protection are present in places with limited statehood and rule of law, where standards that are in force may in fact be unenforceable. Political and economic inequalities between states and regions also create information asymmetries which reinforce and exacerbate these structural barriers to protection. Thus even if the first set of barriers can be overcome, i.e. if LMICs behave like HICs in terms of passing and enforcing data protection standards, the results will not be comparable since they often do not have the same power to rule and enforce, notably with regard to the transnational use of data.

The establishment of clear rules for privacy and data protection worldwide is further complicated by the fact that assumptions about LMIC data are different in various dimensions: the cost-benefit analysis can be skewed by factors such as poverty or crises, making purpose limitation and anonymisation less of a priority; big data is being mythologised as the answer to intractable problems, leading international institutions to advocate for data maximisation; and corporations become more important actors in data flows where states are weak or have few resources, leading to a merging of public and private sector functions with regard to governance and the idea that regulating data flows will harm basic governance.

This powerful discourse about the potential benefits of big data analytical methods is a greater barrier to extending privacy as a fundamental right than it might at first seem. Nissenbaum (2009) argues that the fuzziness of the borders and the essential nature of privacy is its greatest strength, since it provides adaptability and demands ongoing debate. Yet that fuzziness will inevitably make it hard to set the right to privacy against simple statements about how data mining can 'solve' problems such as disease, lack of education, poor sanitation and food insecurity.[13]

Despite the extent of the problem and the multiplicity of ways by which it is configured, HICs with established data regulation regimes may have a particular (and time-limited) chance to affect the fairness of data protection practices worldwide. I have argued above that processes of establishing data protection regulation in LMICs are likely to be strongly influenced by the need for compatibility with trading partners rather than a strong domestic call for a fundamental right to privacy, as was the case in the US and many EU states. Given the inevitable imbalance in power and influence between these regions and their LMIC trading partners, HIC regulatory standards and practices are likely to have disproportionate influence over LMICs in terms of setting the parameters for data use and sharing – but also in terms of offering protection to citizens in regions where HIC corporations are operating.

The most effective way to begin a global dialogue about privacy that is based on the idea of privacy as a fundamental right rather than as a transaction or a cost-benefit analysis, then, may be to reframe LMIC privacy from the international perspective as a regulatory problem that HICs can influence, rather than as a challenge that LMICs must overcome. If the question becomes one of limiting HIC-based corporations' ability to misuse data from LMIC technology users, rather than how LMICs should frame privacy for their own citizens, it becomes both more manageable and more easily related to existing models of regulation such as those set out in section 4. If such a debate can take place, it may drive the evolution of data governance toward and one that does not rely solely on country-level judicial apparatuses to resolve what are fundamentally international asymmetries.

The big data era may come to be seen as the starting point for digital privacy as a global debate. As power through data evolves and replicates itself across sectors and regions, the power to identify and categorise people and to modulate their behaviour is

shifting from the auspices of states to those of international corporations and institutions, with far-reaching implications for individual citizens, and particularly those in countries without enforceable privacy or data protection provisions. On the other hand, however, this extension of the power to make people visible through data may be seen as so extreme that it could also constitute a moment where inequalities can be seen and mitigated. As those outside the HIC technology and regulatory bubble make themselves legible by joining the global technological commons, the next step is to extend to a global context the discussion about how big data analytics affects social existence. It is becoming possible to take into account a diverse range of discourses on data and privacy and to consider how the use of digital technologies makes *everyone* identifiable and legible, in various dimensions and to various authorities and institutions, and how the unrestricted processing and reuse of data reproduces power asymmetries. Paradoxically, however, the most effective way to respond to this imbalance may be on the local level in HICs, by reworking basic local tools as instruments for international influence.

REFERENCES

Banjul Charter (1981) African Charter on Human and Peoples' Rights. Accessed 13.8.2014 at http://www.achpr.org/files/instruments/achpr/banjul_charter.pdf.

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

Borgesius, F.J. (2013) 'Consent to behavioural targeting in European law - What are the policy implications of insights from behavioural economics?', Amsterdam Law School Legal Studies Research Paper No. 2013-43.

Bradford, A. (2012) 'The Brussels Effect'. *Northwestern University Law Review*, 107(1).

Business Standard (2014) Modi backs UIDAI, seeks accelerated DBT rollout. Accessed 29.7.2014 at http://www.business-standard.com/article/economy-policy/modi-backs-uidai-seeks-accelerated-dbt-rollout-114070500756_1.html.

Calo, M.R. (2012) 'Against Notice Skepticism in Privacy (and Elsewhere)', 87 *Notre Dame Law Review*, 1027, 1033.

Castells, M. (1996) *The Rise of the Network Society*. New York: Wiley-Blackwell.

Cavoukian, A., and Castro, D. (2014) 'Big Data and Innovation, Setting the Record Straight: De-identification Does Work', The Information Technology and Innovation Foundation. Accessed online 8.4.15 at https://www.ipc.on.ca/images/Resources/pbd-de-identification_ITIF.pdf.

Cohen, J. E. (2012) *Configuring the networked self: Law, code, and the play of everyday practice*. Yale University Press.

———. (2013) 'What Privacy is For', 126 *Harvard Law Review* 1904.

Deleuze, G. (1992) 'Postscript on the Societies of Control', *October*, vol. 59. pp. 3-7.

Donovan, K.P and Martin, A.Ks. (2014) 'The Rise of African SIM Registration: the Emerging Dynamics of Regulatory Change', *First Monday*, 19 (2-3).

Du Plessis, W. J. (2011) 'African indigenous land rights in a private ownership paradigm', PER: *Potchefstroomse Elektroniese Regsblad*, 14(7), 45-69.

EU 95/46/EC (1995). European Union Data Protection Directive. Accessed online 4.8.14 at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML.

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

Forbes (2014) 'The apparent MPesa monopoly may be set to crumble'. Accessed 28.7.14 at http://fortune.com/2014/06/27/m-pesa-kenya-mobile-payments-competition/.

Gilliom, J. (2001) *Overseers of the poor: Surveillance, resistance, and the limits of privacy*. Chigago: University of Chicago Press.

Global Pulse (2013) United Nations Global Pulse Information Sheet. Available at: www.unglobalpulse.org.

Green, D. (2014) Big Data and Development: Upsides, downsides and a lot of questions. Accessed 4.8.2014 at: http://oxfamblogs.org/fp2p/what-is-the-future-impact-of-big-data/.

Greenleaf, G. (2010) 'India's national id system: Danger grows in a privacy vacuum', *Computer Law & Security Review*, 26(5), 479-491.

———. (2013) 'Sheherezade and the 101 Data Privacy Laws: Origins, Significance and Global Trajectories', *Journal of Law, Information & Science*, 23(4).

GSMA (2011) Mobile privacy principles. Accessed online at: http://www.gsma.com/publicpolicy/wp-content/uploads/2012/03/gsmaprivacyprinciples2012.pdf.

Hardjono, T.; Deegan, P. and Clippinger, J.H. (2014).'Social Use Cases for the ID3 Open Mustard Seed Platform', *Technology and Society Magazine*, IEEE , 33(3), pp.48,54.

Hildebrandt, M. (2013) 'Slaves to Big Data. Or Are We?', Idp. *Revista De Internet, Derecho y Política* 16.

Horrell, M. (Ed.). (1960) *The" Pass Laws."* (Vol. 7). SA Institute of Race Relations.

Irion, K. and Luchetta, G. (2013) Online personal data processing and EU data protection reform. CEPS Digital Forum.

ITU (2013) The world in 2013. International Telecommunications Union. Accessed 4.8.2014 at http://www.unapcict.org/ecohub/the-world-in-2013-ict-facts-and-figures

Jernigan, C., and Mistree, B. (2009) 'Gaydar: Facebook friendships expose sexual orientation', *First Monday*, 14(10).

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

Kahn, E. J. (1966) *The Separated People: A Look at Contemporary South Africa*. New York: WW Norton & Company.

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

Miraftab, F. (2004).'Public-Private Partnerships The Trojan Horse of Neoliberal Development?', *Journal of Planning Education and Research*, 24(1), 89-101.

New India Express (2014). Aadhaar Data Minefield Threatens to Blow Up in Government's Face. Accessed 28.7.14 at http://www.newindianexpress.com/thesundaystandard/Aadhaar-Data-Minefield-Threatens-to-Blow-Up-in-Government%E2%80%99s-Face/2014/06/08/article2268540.ece

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Nyst, C. (2013) 'The road to surveillance is paved with good intentions – and warning signs. Poverty Matters Blog', The Guardian Online. Accessed 4.8.2014 at http://www.theguardian.com/global-development/poverty-matters/2013/nov/12/surveillance-aid-iris-scanning-gps-tracking.

McCallister, E., Grance, T. and Scarfone, K. (2010) Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). National Institute of Standards and Technology Special Publication 800-122.

oAfrica.com (2013) Huawei going strong in Africa. Accessed 28.7.2014 at http://www.oafrica.com/mobile/huawei-going-strong-in-africa/.

Press Association (2014).European court to rule on allegations Facebook passes personal data to NSA. The Guardian, accessed 29.7.14 at http://www.theguardian.com/world/2014/jun/18/facebook-personal-data-nsa

Privacy International (2013) International agreement reached controlling export of mass and intrusive surveillance technology. Press release. Accessed 29.7.14 at http://www.scoop.co.nz/stories/WO1312/S00123/international-agreement-controls-export-of-surveillance-tech.htm

Puschmann, C. and Burgess, J. (2014) 'Metaphors of big data', *International Journal of Communication*, 8, 1690-1709.

Safaricom (2014) MPesa timeline. Accessed 28.7.14 at http://www.safaricom.co.ke/mpesa_timeline/timeline.html

Schermer, B.W. (2011) 'The limits of privacy in automated profiling and data mining', *Computer Law & Security Review*, 27(1), 45-52.

Schroeder, R. (2014) 'Big Data: Towards a More Scientific Social Science and Humanities, in Mark Graham and William H Dutton (eds.), Society and the Internet: How Networks of Information are Changing our Lives. Oxford: OUP.

Scott, J. C. (1998) *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press.

Sharad, K. and Danezis, G. (2013) 'De-anonymizing D4D Datasets', The 13th Privacy Enhancing Technologies Symposium. July 10–12, 2013, Bloomington, Indiana, USA.

Solove, D. J. (2006) 'A taxonomy of privacy', *University of Pennsylvania Law Review*, 477-564.

Srinivasan, J. and Johri, A. (2014). "The role of data in aligning the 'unique identity' infrastructure in India", In Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & Social Computing (pp. 697-709). ACM.

Taylor, L. (2014) 'No Place to Hide? The Ethics and Analytics of Tracking Mobility Using Mobile Phone Data', Border Criminologies blog, University of Oxford. Accessed 4.8.2014 at http://bordercriminologies.law.ox.ac.uk/tag/linnet-taylor/.

Taylor, L. and Schroeder, R. (2014) 'Is bigger better? The emergence of big data as a tool for international development policy', *Geojournal*, online first.

TechMtaa (2012) "Outcry Over Claims That M-Pesa Details Are Used 'To Register' Party Members". Accessed 28.7.14 at http://www.techmtaa.com/2012/01/16/outcry-over-claims-that-m-pesa-details-are-used-to-register-party-members/.

Warren and Brandeis (1890) 'The Right To Privacy', *Harvard Law Review* 4 (193).

White and Case (2012) Data Protection and Privacy in China. Briefing, accessed 4.8.2014 at http://www.whitecase.com/files/Publication/58829ab4-e10d-4371-b722-74681b4ac7e6/Presentation/PublicationAttachment/07ca803c-6ce8-49ae-8b4d-622557551990/alert-Data-Protection-and-Privacy-in%20China-March-2012.pdf.

World Wide Web Foundation (2014) Open Data in Developing Countries: Different models, new approaches. Accessed 4.8.2014 at http://webfoundation.org/2014/07/open-data-in-developing-countries-report-results/.

Yao-Huai, L. (2005) 'Privacy and data privacy issues in contemporary China', *Ethics and Information Technology*, 7(1), 7-15.

NOTES

[1] Article 29 Data Protection Working Party, "Opinion 03/2013 on purpose limitation," April 2, 2013, http://ec.europa.eu/justice/dataprotection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf, p. 4.

[2] Robert Kirkpatrick, interview with *Global Observatory*, 5/11/2012. Accessed online 17/7/2014 at http://theglobalobservatory.org/interviews/377-robert-kirkpatrick-director-of-un-global-pulse-on-the-value-of-big-data.html.

[3] For instance, the role of corporate data-processing strategies in the 2012 US presidential election: see http://www.nytimes.com/2013/06/23/magazine/the-obama-campaigns-digital-masterminds-cash-in.html?pagewanted=all&_r=0.

[4] I use the World Bank's definitions grouping countries, see: http://data.worldbank.org/about/country-classifications, where LMICs have incomes of $1,036 - $12,616 per capita and high income countries (HICS) above that threshhold. My particular focus is the low- and lower-middle-income countries, with an upper threshhold of $4,085 per capita, which includes India and most of Africa.

[5] For example, see http://www.bloombergview.com/articles/2013-06-10/snowden-is-in-hong-kong-chinese-don-t-care-.

[6] IBM research makes this claim for its 'Project Lucy', involving the Watson supercomputer (http://www.research.ibm.com/labs/africa/project-lucy.shtml).

[7] Presentation, WS 203 Big data: promoting development and safeguarding privacy. 8th Internet Governance Forum, Bali, October 22-25 2013.

[8] For a clear example of this belief, see IBM's Watson project in Kenya, which aims to solve Africa's grand challenges using big data analytics (http://www.economist.com/blogs/baobab/2013/11/ibm-africa).

[9] Jack Walsh quote from IGF 2013.

[10] One exception to this is the Indian government's move to regulate the Unique Identification Authority after the 2014 change in leading party (Business Standard 2014).

[11] The discourse on development as a unitary issue lending itself to engineering solutions is common to many big data projects, for example the 'Big Data for Social Good' project at Harvard: http://www.hsph.harvard.edu/ess/.

[12] This was the type of evolution posited by Sunil Abraham, Director of India's Centre for Internet and Society, in a speech at the Internet Governance Forum (Bali, 22-25 October 2013).

[13] See IBM's 'Lucy' project for example: http://www.research.ibm.com/labs/africa/project-lucy.shtml.