Geoforum 64 (2015) 229–237

Contents lists available at ScienceDirect

Geoforum

journal homepage: www.elsevier.com/locate/geoforum

In the name of Development: Power, profit and the datafication of the global South

Linnet Taylor [a,*], Dennis Broeders [b]

[a] University of Amsterdam, The Netherlands
[b] Erasmus University Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 15 October 2014
Received in revised form 2 July 2015
Accepted 3 July 2015
Available online 9 July 2015

Keywords:
Big data
Power
Development
Public private partnerships
Surveillance
Datafication

ABSTRACT

We examine the current 'datafication' process underway in low- and middle-income countries (LMICs), and the power shifts it is creating in the field of international development. The use of new communications and database technologies in LMICs is generating 'big data' (for example from the use of mobile phones, mobile-based financial services and the internet) which is collected and processed by corporations. When shared, these data are also becoming a potentially valuable resource for development research and policy. With these new sources of data, new power structures are emerging within the field of development. We identify two trends in particular, illustrating them with examples: first, the empowerment of public–private partnerships around datafication in LMICs and the consequently growing agency of corporations as development actors. Second, the way commercially generated big data is becoming the foundation for country-level 'data doubles', i.e. digital representations of social phenomena and/or territories that are created in parallel with, and sometimes in lieu of, national data and statistics. We explore the resulting shift from legibility (Scott, 1998) to visibility, and the implications of seeing development interventions as a byproduct of larger-scale processes of informational capitalism.

1. Introduction

There is a process of 'datafication' (Mayer-Schönberger and Cukier, 2013) underway in low- and middle-income countries (LMICs),[1] where the use of new communications and database technologies in LMICs is generating digital data that is machine-readable and computationally manipulable, particularly for 'big data' analytics. These born-digital datasets are of unprecedented size and detail, especially compared to the statistical records previously available on lower-income countries (Jerven, 2013). In contrast to traditional state survey data, however, these data are generated, collected and processed under the auspices of private-sector corporations and are shared, often on a pro-bono basis, at the level of international academic research institutions or development actors such as the UN. Where previously development donors (governments or international NGOs) worked with LMICs' own statistical apparatuses to generate population data, it is becoming increasingly possible and cost-efficient for donors to turn to corporations for consumer-generated data that can proxy for traditional household surveys and other statistical products (Taylor and Schroeder, 2014). The discourse on big data as a resource for development (World Economic Forum, 2012; Global Pulse, 2012; Taylor and Schroeder, 2014) indicates that a shift is underway from the predominance of state-collected data as a way of defining identities and sorting and categorising individuals, groups and whole societies to a big-data model where data is primarily collected and processed by corporations and only secondarily accessed by governmental authorities.

The central question addressed in this paper is how datafication is influencing the way that LMIC populations are made legible in the context of development, and what this means for power dynamics amongst development actors. We examine a power shift from the traditional collector and user of statistics – the state – to a messier, more distributed landscape of governance where power accrues to those who hold the most data. This power shift has its roots in the larger neoliberal trend in governance worldwide – what Cohen (2013: 1928) has termed the 'new governance,' dominated by public–private partnerships engaged in 'informational capitalism' (2013: 1912, following Castells (1996)), a system where 'information flows in circuits that serve the interests of powerful

* Corresponding author.
E-mail address: l.e.m.taylor@uva.nl (L. Taylor).

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

http://dx.doi.org/10.1016/j.geoforum.2015.07.002
0016-7185/

entities, both private and public' (2013: 1916) and where those who trade in information – primarily corporations – are able to modulate people's behaviour, activities, and relationship to the state. She argues that the creation of surveillance infrastructures are an inevitable part of governance through informationalism, and the possibility of surveillance is therefore pervasive. We argue that 'data-driven development' is characterised by the same features as informational capitalism, and results in increased visibility for the populations of lower-income countries – though not necessarily in greater governability or representation.

This shift towards a combination of datafication and privatisation is still in its early stages and the evidence is not yet available to draw conclusions about its medium or longer-term impacts. It is potentially wide-ranging, however, and has powerful implications for how LMICs will be understood by the development field as economic and political actors in the coming years. So far, little research has been conducted on the political implications of the production of big data about lower-income countries (exceptions include Thomas (2014), Burns (2014) and Taylor and Schroeder (2014)). We therefore assess the longer-term risks of the secondary use of such data under the rubric of economic or human development, based on illustrative examples of current data collection by corporations and public–private partnerships in LMICs and aim to contribute towards a research agenda on this issue by drawing out and illustrating key trends and their potential repercussions.

We define two key trends using illustrative examples: first, the increasing empowerment of private sector actors in the field of international development due to their ownership of data, and second, the emergence of development/surveillance assemblages with the potential for ongoing monitoring of population dynamics and people's activities. The first trend involves leverage and opportunities for corporate actors who are proprietors of vast quantities of data on citizens of developing countries, whilst states continue to prioritise traditional survey data such as censuses which convey a different type of detail and allow for different types of sorting and monitoring. We look at how the deputisation of technology corporations by states through public–private partnerships (PPPs) may lead to the merging of these different analytical perspectives. We then frame the second trend, the emergence of country data doubles and various forms of shadow mapping – in relation to these partnerships.

This paper is based on 60 interviews on the use of big data in development policy and planning, along with two years' research (2013–2015) on the interface of data science and international development policy. The research also involved attending international conferences, workshops and public discussions hosted by large institutional actors, and by gathering knowledge from mailing lists and other online discussions. This project was partly funded by the Alfred P. Sloan Foundation at the Oxford Internet Institute. 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. 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.

2. Power through LMIC data

2.1. Legibility, visibility and emergent power structures

Scott's (1998) famous book *Seeing Like a State* could instead have been titled 'reading like a state' since the registrations upon which the modern bureaucratic state was built, made society – in Scott's own words – 'legible'. 'Reading like a state' in the era of big data, however, increasingly involves remotely performing data analytics to make populations visible. This constitutes a step beyond – or away from – Scott's formulation because such visibility creates power over data subjects via volume of data rather than accuracy and detail. Unlike legibility, which depends on data that people are aware of providing, the data that provides visibility is often observed, not volunteered (Hildebrandt, 2013), derived as a byproduct of technology use rather than collected by authorities through survey methods. Unlike the data that provides legibility, these data are of unknown reliability because they reflect, in Shearmur's words (2015), not populations but 'users and markets' and are therefore biased towards those with connectivity. This unknown bias is compounded because the visibility such data offers is created by data scientists, not social scientists, and frequently lacks contextual information to explain what is being seen (Taylor and Schroeder, 2014). The gap in terms of distance – but also cultural dissimilarity - between the processors and end users of LMIC big data and those living in these countries may also increase power asymmetries. This suggests that the power to make visible is different from the power to make legible. Legibility increases governability (in Scott's formulation), but visibility offers the power to influence and intervene to a wider, more distributed set of actors: the corporations who gather and analyse the data, plus whoever they choose to share it with (or can capture it through other means), who may be state actors, international development institutions, or other corporate partners.

Traditional visualisations such as mapping are being transformed by spatial big data analytics. GPS-derived big data from mobile phones in particular lends itself to real-time, highly detailed mapping of mobilities (Taylor, 2015a), and sensor data offer new possibilities for understanding urban space in particular (Kitchin, 2014). However, this big data is mainly created and accessed within government–corporate networks (Crampton et al., 2014), so that big data is a tool of the powerful in the corporate and governmental spheres who can afford to collect, merge and analyse it. In this way, little has changed since Harley's assertion (1989) that maps are essentially political despite the claim of scientific objectivity put forward by the profession of cartographers. It also reinforces and broadens Taylor and Johnston's (1995: 58) warning to geographers that by using digital data generated by state apparatuses, they are 'creating the state's geography' and determining 'what is included and what is not'. Today, a similarly critical approach is needed in relation to corporations that collect and own data, and determine how they may be made accessible. What kind of geography is emerging though the corporate-led Data for Development?

Power is thus exerted through the generation of data, control of access to data, the merging of various data, but also through their analysis and interpretation. At each of these points, however, the private sector's role is increasingly important but also increasingly hard to pin down. Bowker and Star (2000) in their seminal work on classification remarked that information architectures (and the politics that they contain) are hard to analyse because 'Good usable systems disappear almost by definition. The easier they are to use, the harder they are to see. As well, most of the time, the bigger they are, the harder they are to see' (Bowker and Star, 2000: 33). At this moment, big data analytics constitute the biggest usable system around and the corporate and governmental enthusiasm for its potential does not motivate many to analyse what lies under the hood. Similarly, Schneier (2015: 30) points out that data collection in HICs has become so engrained in modern life that it fades into the background for the individual user. Moreover, most of the attention that has been paid to projects under the rubric of Data for Development focuses on the perceived and expected benefits at the expense of the epistemological and power implications of the data analytics used.