Development Informatics

Working Paper Series

The Development Informatics working paper series discusses the broad issues surrounding digital data, information, knowledge, information systems, and information and communication technologies in the process of socio-economic development

An Applied Data Justice Framework*: Analysing Datafication and Marginalised Communities in Cities of the Global South*

RICHARD HEEKS & SATYARUPA SHEKHAR

2018

View/Download from:
http://www.gdi.manchester.ac.uk/research/publications/di/

Educators' Guide from:
http://www.gdi.manchester.ac.uk/research/publications/di/educators-guide/

Table of Contents

An Applied Data Justice Framework*:*
Analysing Datafication and Marginalised Communities in Cities of the Global South

Richard Heeks	**&**	**Satyarupa Shekhar**
Centre for Development Informatics, University of Manchester, UK		Citizen Consumer and Civic Action Group, Chennai, India

2018

Abstract

Rapid recent growth in the role of data within international development has meant analysis of this phenomenon has been lagging; particularly, analysis of broader impacts of real-world initiatives. Addressing this gap through a focus on data's increasing presence in urban development, this paper makes two contributions. First – drawing from the emerging literature on "data justice" – it presents an explicit, systematic and comprehensive new framework that can be used for analysis of datafication. Second, it applies the framework to four initiatives in cities of the global South that capture and visualise new data about marginalised communities: residents living in slums and other informal settlements. Analysing across procedural, rights, instrumental and structural dimensions, it finds these initiatives deliver real incremental gains for their target communities. But it is external actors and wealthier communities that gain more; thus increasing relative inequality.

A. Introduction

Datafication – the presence, use and impact of data in social processes – is a growing phenomenon worldwide, including in the domain of international development (Taylor & Broeders 2015). This growth has seen both hopes and concerns for the developmental impact of new data streams (Spratt & Baker 2015). But, because of the novelty of the phenomenon, writing has sometimes had to be based on overviews or extrapolations from pilot projects or experiences in the global North (*ibid.*, Lokanathan 2017). There have been relatively few analyses to date of real-world initiatives, and the field is still assessing analytical approaches to use (Sengupta et al 2017).

This paper thus aims to do two things. First, to introduce a conceptual framework drawn from the sub-field of data justice that can be used for analysis of data initiatives. Second, to utilise that framework to analyse a set of data initiatives within the domain of urban development; in particular seeking to understand the full breadth of impact of such initiatives on those who are already marginalised within the physical city.

There follows next a review of datafication, particularly in relation to urban development, and presentation of the data justice analytical framework. Then the methods used for this paper – study of urban data initiatives in Chennai, Nairobi, Pune and Solo through primary and secondary sources – are explained. Findings and conclusions follow.

B. Datafication, (Urban) Development and Data Justice

Data has always been part of international development as it – or at least its processed and utilised form, information – is essential and integral to all decisions, including development-related decisions. But the advent of digital technologies and their significant diffusion in developing countries during the 21st century have led to a step change. The increasing datafication of development can be described in terms of a growing volume, velocity, variety and visibility of data (Heeks 2018). This increasing presence of data within development is part-encouraging, part-driven-by greater use of new forms and streams of data in decision-making; in turn leading to emergent developmental impacts associated with this datafication (*ibid.*).

This broader picture is instantiated in the various development domains, including the particular focus here: urban development, with the expanding role and importance of data recognised within the New Urban Agenda that is intended to guide future development (UN-Habitat 2017). New forms and flows of data are especially associated with growth of the "smart city" agenda in the global South (Shekhar 2016) and with the affordances of digital technologies: mapping cities by remote sensing, gathering environmental and other data via ground-based sensors, creation of "data twins" of urban areas such as mobility maps created by tracking mobile phone use (Lokanathan et al 2016, Willis 2017).

These initiatives come with much promise around the benefits they will bring to urban planning and development; improving its efficiency by enabling decisions and results to occur more quickly and/or more cheaply; and improving its effectiveness by enabling better decisions and results: better-informed, more objective, more transparent, better-able to meet citizens' needs (van Veenstra et al 2014, Samarajiva et al 2015). But alongside this have been concerns about datafication of the city (Townsend 2013, Kitchin 2014, Taylor & Richter 2015, Baud 2016): procedurally that initiatives are not being implemented right; instrumentally that the promised results are not being achieved; and critically that there are problematic distributive impacts. Main critical concerns are that urban datafication is associated with growing inequality; especially, in developing countries, with the exclusion or adverse incorporation of those already marginalised within the physical city, such as those living in slums and other forms of informal settlement (Donovan 2012, Pfeffer & Verrest 2016).

Given the extent of urban inequality – 30% of the urban population in developing countries lives in slums (UN-Habitat 2016) – and the history of this population being excluded from or marginalised by new initiatives, some development actors have begun what we may call pro-equity data initiatives (PEDIs). Though coming with other labels attached, an essence of these activities has been creation of new dataflows and datasets by and/or about and/or for slum communities. Examples include community wi-fi that enables new dataflows within, into and out of slum communities; open data / right-to-information that enables new dataflows into slum communities; and community mapping that creates new datasets about slum communities (Chakraborty et al 2015, Willis 2017).

But how should these (and other) data initiatives be understood and evaluated, given recent recognition of a lack of evaluative frameworks to analyse datafication and development

(Hagen 2017, Sengupta et al 2017)? One approach would follow descriptors used for the datafication concerns expressed above: "injustice" (Townsend 2013), "ethical concerns" (Kitchin 2014), "ethical challenges" (Taylor & Richter 2015). This would suggest analysis using ideas within the emerging literature on "data justice": "the specification and pursuit of ethical standards for data-related resources, processes and structures" (Heeks 2017a); a small body of work within critical data studies echoing our specific interests around equity.

This has been conceptualised from a capabilities perspective (Taylor 2017, Heeks & Renken 2018) but we wanted to take a broader view of data justice that specifically encompassed three things highlighted in recent literature. First, the concerns about urban datafication expressed earlier: procedural, instrumental, critical. Second, a specific focus on data and data praxis and data rights given the understanding from critical data studies that any social injustices may relate more to forms and flows and use of data than to specifics of the technologies that carry and process and display that data (Taylor & Broeders 2015, Neff et al 2017). Third, the argument of other critical data studies literature that broader social structure must be incorporated into any analysis because of its role in shaping data-related outcomes (Dalton & Thatcher 2014, Kitchin & Lauriault 2018).

The model developed from these principles is summarised in Figure 1 (adapted from Heeks 2017b, which should be referred to for further details of its development) and examines five dimensions of data justice:

- Procedural: fairness in the way in which data is handled. This handling is understood in terms of the information value chain: the steps by which data is transformed into developmental results (see Figure 2).
- Instrumental: fairness in the results of data being used.
- Rights-based: adherence to basic data rights such as representation, privacy, access and ownership.
- Structural: the degree to which the interests and power in wider society support fair outcomes in other forms of data justice.
- Distributive: an overarching dimension relating to the (in)equality of data-related outcomes that can be applied to each of the other dimensions of data justice.
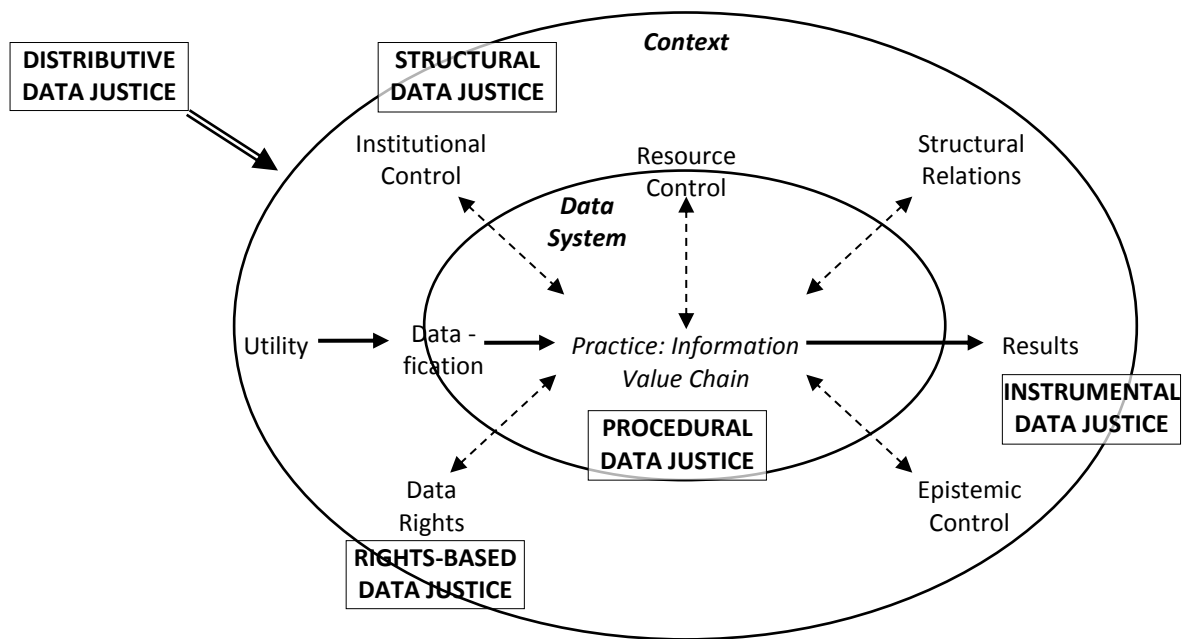
Figure 1: Conceptual Model of Data Justice (adapted from Heeks 2017b)

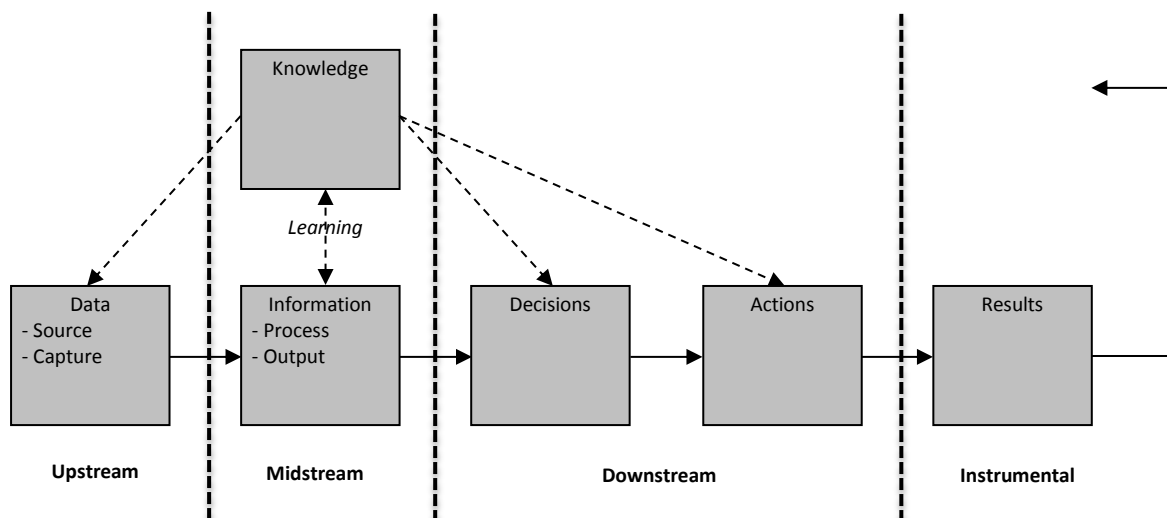**Figure 2: The Information Value Chain (adapted from Heeks 2017b)**[1]

Further explanation of the two models will follow below through their use as analytical frameworks for evaluation of four urban PEDIs, following an outline of the initiatives and methods used for evidence-gathering.

---

[1] Knowledge links to all steps of the chain and therefore is not associated with any individual stage: up-, mid- or downstream.

C. Case Background and Methods

Among the possible types of pro-equity data initiative, we chose those involving community mapping, which aims to counter the relative invisibility of (i.e. lack of data about) marginalised communities by gathering, visualising and utilising new data on locations, assets and issues within those communities. Mapping was chosen because it is quite well-established and widespread as a PEDI in cities of the global South (Patel & Baptist 2012, Panek & Sobotova 2015). While the analysis here speaks to broader trends of datafication, this means there is a ready direct literature and constituency of relevance.

The four initiatives chosen for analysis are:
- Map Kibera (MK): begun in 2009 by two Americans, this used local young people to gather data on Kibera, a slum roughly estimated at 250,000 people located in South-West Nairobi, and then mapped that data. Via creation of social enterprise GroundTruth Initiative, expertise from Kibera has subsequently been applied to other slum areas in Africa and the Middle East.
- Our Pune Our Budget (OPOB): begun in 2006 by local NGO, the Centre for Environment Education (CEE), this sought to bring greater objectivity and fairness into the city's budgeting process, as a foundation for which, it organised a city-wide data-gathering and mapping exercise.
- Solo Kota Kita (SKK): begun in 2010 Kota Kita is an Indonesian NGO based in the city of Solo (official name, Surakarta). Its Solo Kota Kita project sought to enable citizen participation in the participatory budgeting process initiated by mayor (currently Indonesian President) Joko Widodo. It did this by gathering and mapping data across all city neighbourhoods. Kota Kita has subsequently rolled out its methodology to other cities across Asia.
- Transparent Chennai (TC): begun in 2009, this involved a series of projects to gather and map data on issues within the city's marginalised communities including lack of public services, homelessness, and informal livelihoods (specifically of waste-pickers). In 2014, Transparent Chennai was rolled into a wider initiative, the Transparent Cities Network – housed in local NGO Citizen Consumer and Civic Action Group – that focuses on data and the urban poor.

These were chosen because they were relatively long-standing PEDIs[2], which – unlike many initiatives – moved well beyond pilot stage; because there existed some secondary literature for each; and because they operated in cities with substantial slum populations: 50% of the city population in the case of Nairobi (Kovacic & Lundine 2013), 40% in Pune (Jobst & Malherbe 2017), at least 8% in Solo (Obermayr 2017), and 28% in Chennai (Krishnamurthy & Desouza 2015).

The findings below are drawn from four sources which were combined and then analysed through the thematic lens of the data justice model presented above. First, nine interviews

---

[2] MK, OPOB and TC are 'targeted' pro-equity initiatives that specifically sought to reduce marginalisation of slum communities; though in OPOB's case this was seen to require city-wide data. SKK is an 'inclusive' pro-equity initiative in seeking to ensure those in informal housing were included within a wider pro-equity goal of "levelling the playing field" between citizens and government so that citizens could shape urban budgeting decisions.

with senior figures in the data intermediary organisations (MK, CEE, SKK, TC) associated with each of the four projects, structured around the concepts of data justice[3]. Second, evaluation documents from the projects which are reflective, to some degree self-critical, and sometimes themselves based on primary fieldwork (e.g. Menon 2013, Padmanabhan et al 2013, Rifai et al 2016, Hagen 2017). Third, independent secondary sources of research specifically on these projects, most of which are based on primary fieldwork (e.g. Berdou 2011, Grillos 2017, Jobst & Malherbe 2017). Fourth, broader independent sources that incorporate analysis of these projects among a number of others (e.g. Donovan 2012, Haklay 2013, Shkabatur 2014, Baud 2016).

C1. Background: Data Injustice and its Consequences

Before analysing the cases in terms of data justice, we first outline the rationale behind the initiatives: the data injustices and consequences thereof which obtained prior to the PEDIs being undertaken.

The communities studied suffer a range of injustices but of specific attention here was the data rights injustice relating to representation: that the physical marginalisation of these communities was mirrored by a virtual marginalisation within urban datasets. The communities were invisible or poorly visible: not merely that the areas themselves were unmapped as locations but that data on the assets, services, voices and livelihoods of citizen were absent (iMK1b, iOPOB2, iSKK). Data had sometimes been gathered – by NGOs and academics in Kibera, by government in Chennai and Pune – but it fell short on the five "OCARA" measures of data quality: openness, completeness, accuracy, relevance, appropriateness of presentation (Heeks 2018). Data on slums was generally treated as proprietary, and was inaccessible to the community or to other organisations (iMK2, iTC1, Shekhar & Padmanabhan 2015). Where it might be accessed, the data covered only part of the settlement, or was many years out-of-date, or was held on paper forms and records scattered across many different government departments (iOPOB2, iTC2, Shekhar & Padmanabhan 2015).

This marginalisation within the data twin of the city – an inequality compared to the better representation of more prosperous and formalised settlements (iTC2, Patel & Baptist 2012) – underpinned infrastructural and political marginalisation. The lack and poor quality of slum data "directly results in poor planning and maintenance of public infrastructure, and poor provision of public services" (Shekhar & Padmanabhan 2015:3); for instance, with lack of data on concentrations of population leading new toilets in Chennai to be planned and placed where government had available, accessible land rather than where they were needed (iTC3), or with water points in Kibera poorly maintained due to lack of data on their functioning (iMK1b), and with every type of public service – water, sanitation, roads, lighting, health, education, policing, etc – provided to a very limited degree or poorly provided or simply not provided at all, in part due to lack of data about locations, needs, existing infrastructure, etc (iMK1a, iOPOB1, iSKK, iTC3).

---

[3] These will be referred to below as iMK1a, iMK1b, iMK2, iOPOB1, iOPOB2, iSKK, iTC1, iTC2 and iTC3.

Lack of data also marginalised communities politically, ceding and skewing power within decision-making and service provision to political elites and their interests. With data on neither what the community already had nor on what it needed, there was no basis for engagement with urban planning decisions, nor a basis to challenge the decisions already being made, leading to political exclusion (iOPOB2, iTC2). Those who are data-less in the city – the slum dwellers, the homeless, the waste pickers – could be treated much worse than other citizens – subject to police raids or evictions – with few consequences for those guiding these actions (iTC2, iTC3, Otiso 2002, Padmanabhan et al 2013). Likewise there was a lack of political accountability around provision of public services given the absence of data on provision decisions or on provision itself (iTC1, Menon 2013, Feruglio & Rifai 2017).

D. Findings

As per the model shown in Figure 1, the findings here will be presented in terms of each of the dimensions of data justice in turn, with distributive data justice considered in the final section.

D1. Procedural Data Justice

Because of the sampling or transect approaches used, only a minority of community members could be direct respondents within the four projects. In only one case – Map Kibera – were community members used to capture data (iMK1b). In other cases, data capture was undertaken by students (in Solo and Pune and one Chennai project) or by NGO or local government staff in the other Chennai projects (iOPOB2, iSKK, iTC1, iTC2, iTC3). The young people employed from Kibera were able to upload the data captured on their GPS devices but in this and in all other cases, the "midstream" (see Figure 2) activities of recording data onto a geographic information system, processing it, and visualising – for example printing out as a paper-based map – were generally undertaken by voluntary or paid professionals from outside the community (iMK2, iOPOB2, iSKK, iTC1, Hagen 2011).

The data produced by PEDIs is not always used: sometimes data is gathered without downstream information value chain processes existing that could feed that data into decisions and actions (iMK1b, Shkabatur 2014, Hagen 2017); sometimes those processes exist but they are not accessible, such as government decisions about urban planning or service delivery or evictions for which data from the community is not wanted, or for which officials are unaware of data being available (iOPOB1, iSKK). It may also be used only indirectly: as part of general advocacy, such as pressure on government by NGOs or local media for slum improvements (iOPOB2, iSKK, iTC1).

Where data <u>was</u> directly used, those involved in downstream processes might either be within or outside the community. For the latter, impact was understood mainly in terms of visibility and results, discussed below. Regarding the former then, for the Pune and Transparent Chennai initiatives, it was not the intention that data would be used by the community. For Map Kibera, this was an aspiration but one that was not realised (iMK1a). In all these cases, then, any tracked use of community data was by organisations outside the

community. In Solo, data was fed back to communities in the form of a "mini-atlas" (see Figure 3; a translation of the actual map provided in Bahasa Indonesia). This was present in community meetings held as part of the overall process of participatory budgeting in the city. However, the minority of informal settlements and poorer residents in each neighbourhood plus time, confidence, literacy and other barriers to participation of the poor meant any use of the data for prioritisation decisions was dominated by richer residents (iSKK, Grillos 2017). And these meetings were in any case only the first in a chain of budget decisions that were taken outside the community.

In general then, although all of these were intended as pro-equity data initiatives that would counteract the marginalisation of informal settlements, they were somewhat "extractive" in utilising some community residents as data sources but largely excluding them from all other information value chain processes. Exclusion from up-and midstream processes tended to be on instrumental grounds that they lacked skills or motivation and/or that mapping had to meet a short time-scale. Exclusion from downstream processes tended to be for more systemic or political reasons, such as an absence of mechanisms for involving citizens in service provision or urban development decisions. Exceptions only occurred through deliberate efforts of data intermediary organisations; and took quite some effort to achieve.

It would be normal to decry this limited usage and to seek prioritisation of the 180° approach that extracts data from communities but then feeds it back into those communities so that residents and community-based organisations (CBOs) can improve their decision-making (Sanchez et al 2013). Of course this has merit but Berdou (2011) observes that such prioritisation may be an external imposition. Kiberans themselves prioritised the downstream use of their data not by the community, which may already know itself to some extent, but by outside agencies which do not know the community; believing this was the best way to bring in what they wanted, which was money – via visitors or donor funding or even business investment.

Nor should one ignore the procedural benefits for communities and their members. Indeed, one argument has been that these benefits have tended to be relatively ignored or downplayed compared to the instrumental impact of community data projects (iMK1b, Shkabatur 2014). Even where not directly used for decisions and actions, data has value if fed back into the community e.g. via posters displayed at meetings or on public buildings (Patel & Baptist 2012, Hagen 2017, Jobst & Malherbe 2017). There was an interpretive value: giving residents a feeling of pride and of greater legitimacy as citizens. And there was an epistemic value: making residents aware of issues that need addressing, and giving them knowledge of local services they might previously have been unaware of e.g. the nearby presence of a school, clinic, water point, toilet, etc.
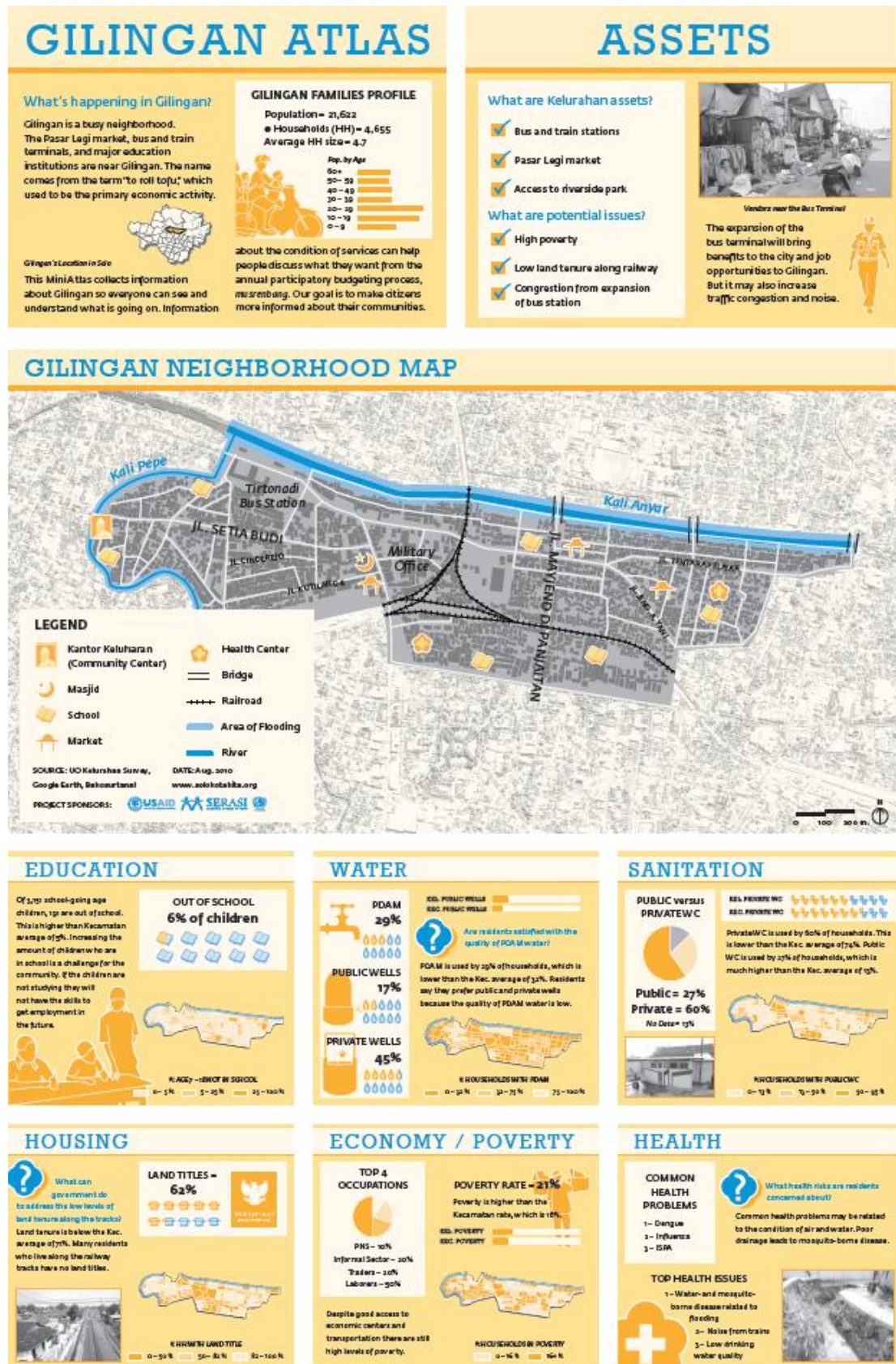
Figure 3: Example Mini-Atlas from Solo Kota Kita (SKK 2010)

This broad but shallow impact of data capture and dissemination ran alongside reports of deeper but much narrower impacts on those who were gathering the data. They gained human capital in the form of technical skills and confidence and greater knowledge of the community, and social capital – both quantum of connections and qualitative gains of greater trust – in the form of a network of contacts in the community, intermediary organisation and more widely (iMK1b, iTC1, Berdou 2011). But, of course, only in the case of Map Kibera were these gains for community members, and the social capital then only arose due to deliberate efforts to send the Kiberan mappers to academic conferences, tech workshops and so forth (Hagen 2017).

There were other process gains but they always lay outside the community: higher-level technical skills developed by the data professionals involved in midstream activities (iMK2, Haklay 2013), and a combination of human and social capital developed by those within the intermediary organisation involved in using the data for advocacy purposes and in encouraging data-utilising activities by government such as budgeting or other urban planning decisions (iMK1a, iSKK, iTC1). While these external stakeholders continued their relation with the informal settlements, these capital accumulations might be applied on behalf of the community. So, also, for the interpretive impact of data on external stakeholders. This was most often expressed as the changed worldview or "slum imaginary" of government officials; from perceiving Solo residents as passive and ignorant to seeing them as active and knowledgeable (iSKK); from seeing Chennai slums as illegitimate to accepting if not ownership rights than at least rights to basic services (iTC1); from seeing Kibera as asset-poor to asset-richer (iMK1b)[4].

D2. Rights-Based Data Justice

Data rights – particularly the right to be fairly represented in datasets and the right to hold some aspects private; but also rights of data access and ownership – were most often discussed in terms of who and what was visible, and to whom.

Informal settlements have been partly- sometimes wholly-invisible within the city; denying them right of representation. All of the PEDIs render these settlements more visible in some way, mapping or otherwise recording their locations, boundaries, buildings, roads, paths; their assets: e.g. clinics, water points, public toilets, schools, businesses, markets, religious buildings, community offices, lighting; and their issues: e.g. health problems, crime hotspots, levels of poverty (iMK1b, iOPOB1, iSKK, iTC1, iTC2, iTC3). An illustrative comparison can be made, for instance, between Kibera's appearance on Google Maps and on OpenStreetMap; the latter used to record some of Map Kibera's data (see Figure 4).

---

[4] Hagen (2017) cites the example of the District Education Officer expecting there to be around 100 schools in Kibera where the data mapped from the community showed her that there were more than 330.
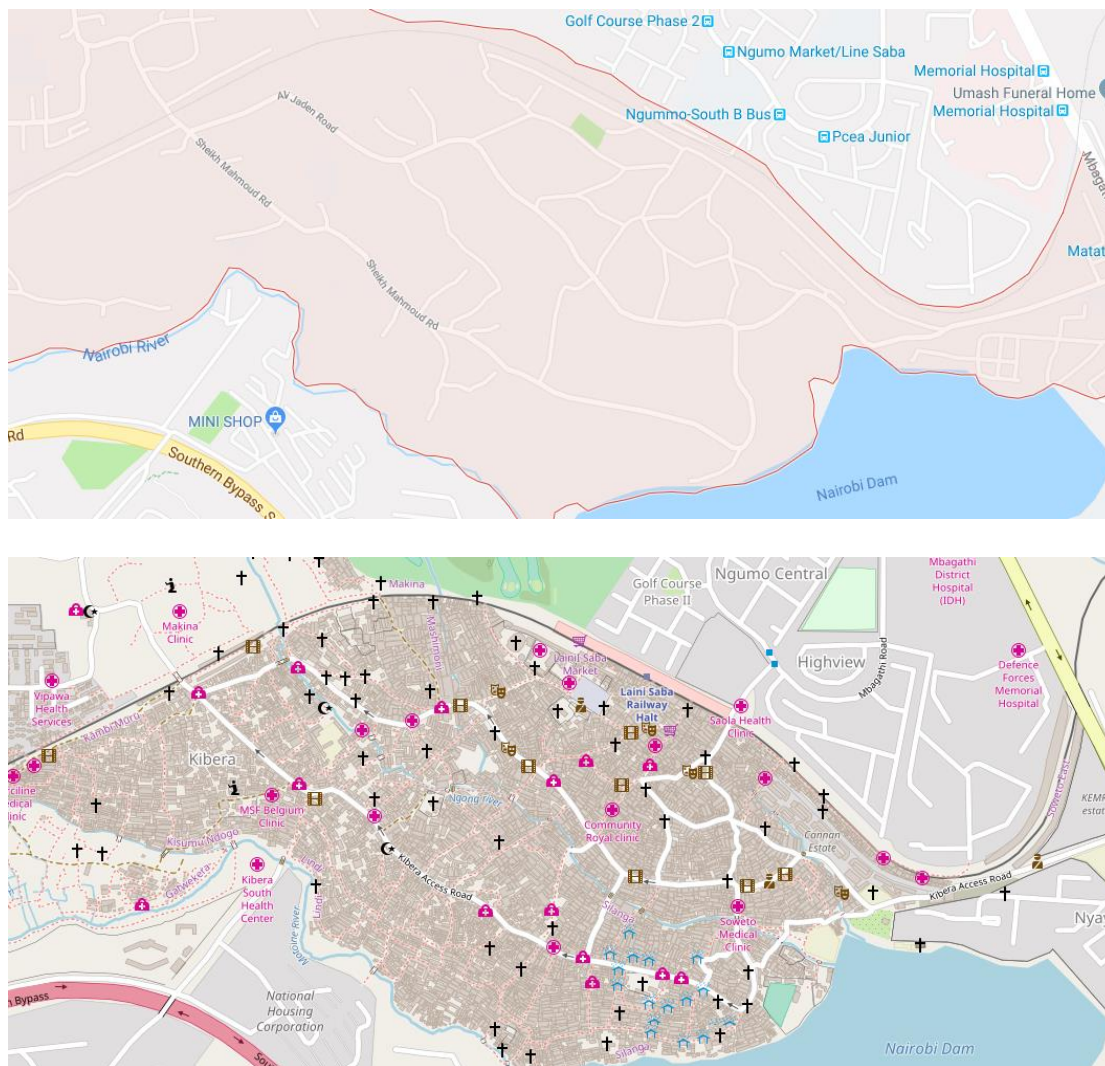
Figure 4: Visible Slums: Google Street Map (top) vs. OpenStreetMap (bottom) views of Kibera

Right of access to these digital representations of the community was skewed. Although growing over time, absence or high cost of mobile let alone fixed-line internet was notable in the communities. As a result, they had to access data in paper-based form; which is more limited in utility (iMK2, iSKK). For external stakeholders, digital access was much easier.

This external visibility – an exercised right to more accurate representation – brought with it various perceived values. As noted above, there were changed perceptions of slums, and some resident felt this gave them some type of legitimacy and would offer them protection, for example from eviction (Padmanabhan et al 2013, Hagen 2017). This wider visibility even impacted some government officials where they felt that open availability of data via the web might attract national or even international attention to their actions (Kovacic & Lundine 2013). And, as discussed in the next section, there was a perceived instrumental value: the more complete, accurate and objective data now available was seen as the basis to replace or to challenge past decision-making processes that were variously regarded as non-existent and/or of poor quality and/or driven by the politics of external stakeholder interests (iMK1b, iOPOB2, iSKK, iTC2).

But this external visibility draws in Scott's (1998) notion of legibility: the process by which states simplify and standardise data about their citizens in order to control them. Legibility disrupts local control over data and knowledge, undermining the power and the value of local understandings within slums (Donovan 2012, Sanchez et al 2013). When data is owned and held within the local community, there is local control; no doubt skewed and imperfect but localised nonetheless. As data becomes more widely circulated – up to the point where web-based digital forms are available to anyone, anywhere with internet connectivity – that local control is lost. These slum areas – all of which now exist virtually in some form of web-based map – are now legible not just to the state but to local and international NGOs, donor agencies, media organisations, academics (iMK1b, iSKK); all of whom can know the slum to some extent and make decisions and actions using data about the slum without the permission or even the knowledge of anyone living there. The right to be represented can thus mean that the right to own and control is ceded to these external agents and to their particular interests and agendas, benevolent or otherwise.

This loss of control and histories of data extraction to the benefit of others but not the community have led to resistance to data capture; with rights to privacy being asserted. Residents in Kibera were often antagonistic to data being captured by video after years of NGOs, movie-makers, tourists, media organisations coming to the area, and making it visible to external audiences for their own rather than the residents' advantage (Hagen 2011, Benequista 2015). Some businesses – schools, pharmacies – in Kibera did not wish to be mapped, fearing their visibility to the state might lead to closure if their location became known and their informal status or activities (e.g. sales of stolen drugs) were then discovered (Berdou 2011). Particular settlements in Chennai refused to participate in data-gathering as they believed drawing attention to their existence and informal status – being under the "gaze of the state" – would increase the likelihood of eviction (iTC1). Transparent Chennai itself had concerns about this – for example, capturing data on issues facing informal waste-pickers in the city but not their location or legal status, in order to protect them from state action (iTC3). If given a choice (which many were not) most slum dwellers had expressed primacy of their right of representation: to be incorporated into urban datasets. But for some this was seen to be in tension with their right to privacy, and for a few who wished to fly under the state's radar the latter was dominant.

While varying across projects, local residents could sometimes shape the balance between rights of representation and privacy, making an input to determination of what data should be captured and, hence conversely, what remained invisible. More broadly, there should be recognition of who and what remain invisible, without a realised right to representation. Transient populations (e.g. labourers or street sellers who come daily into the city to make a living but then leave), the homeless[5] and in some cases those without legal citizenship were excluded from data gathering and hence from representation (Feruglio & Rifai 2017). Qualitative data was lost: what is captured is simplified and deracinated, stripping away the local meanings and histories and relations and all other aspects of local knowledge into which data is always embedded, leaving knowledge-less data that is mobile and readily inserted into external contexts and processes (Hagen 2011).

---

[5] Save one Transparent Chennai project specifically mapping homelessness (iTC1).

If we look at urban decisions, then data can be provided about four elements: background, priorities, process and results. Background data – location, assets, etc – about slum communities was universally provided. The needs and priorities of those communities were sometimes made visible and sometimes not. But other aspects often stayed invisible with community members being granted no right to access: the process by which decisions were made e.g. by government officials; the final decisions; and the extent of implementation of those decisions. As a result, community members were unable to either participate in or monitor the decisions being made about them unless the data intermediary organisations undertook specific actions to counteract this (iTC2, Menon 2013, Feruglio & Rifai 2017).

Finally, we observed what may be called "spotlight and shadow": by rendering some issues more visible, others can be rendered relatively more invisible. This was most notable in the Transparent Chennai case where TC itself and other groups had collected and published data about the resettlement tenement blocks into which some slum residents had been moved following eviction. They saw some limited improvements in those blocks but that this data activity "would also legitimise or validate these tenements and the approach of the government to evict slum dwellers and resettle them in far-flung, poor quality tenements" (iTC1). By highlighting the tenements, data-gathering was seen to adumbrate the eviction process and to encourage an increased rate of tenement building and eviction in the city: representation of one group and issue thus 'unrepresenting' another group and issue. As a result, TC had stopped gathering this type of data.

D3. Instrumental Data Justice

The evidence available is of relatively few results arising from direct decisions and actions by community residents and community-based organisations. These results likely exist – residents using a school or clinic they were previously unaware of; residents attending a community meeting having seen a map of issues – but they were not captured. This reflects the more general point noted above. Users of the open data twin of communities – whether on a public poster or loaded onto a web site – and any resulting actions within the community are untraceable unless they contact or otherwise come to the attention of data intermediaries (Hagen 2017).

In all four cases, government was a downstream presence. Systematic tracking of the value chain through to results seemed, unfortunately, to be rather rare; let alone independent verification. Evidence was thus based more on anecdotes and claims than one would have wished, and more on decisions such as plans than on implemented actions:
• Gathered data – in the form of community "mini-atlases" – was fed into community decision making to set priorities for spending of devolved budgets in Solo. The actual use of the data as information for prioritisation appears to have been somewhat limited (iSKK). Rifai et al (2016:63) describe this process as "challenging", in part because it is management committees that then make final decisions on which projects will be funded. While the end result is community development projects, the link to data is unclear, and analysis shows spending to flow disproportionately to the least-poor areas (Grillos 2017).

- Community data was used in Pune as the basis to set an objective, needs-based budget that took account of the prevalence of poverty and inadequate housing across the city. Unfortunately that budget was then "unset" by local politicians (iOPOB1). Slum areas did see some investment but there was a skew towards middle-class issues (e.g. road building/improvement) and areas: 40% of Pune's population lives in slums but only 10% of the budget was allocated for slum improvement (Menon 2013).
- Data from Map Kibera was used to improve government resource flows into Kibera schools, including involvement of informal school teachers in government-run activities such as a teachers' retreat (Hagen 2017). But the resource flow was, as expected, mainly to the 25% of schools that were government-run, while the extent and sustainability of wider results was unclear.
- Transparent Chennai engaged more with government from the start of the project but the reported outcomes were rather circumscribed: the city built 15 homeless shelters not the 75 that the data had shown to be required (iTC1), or laid on a ceremony to present a few waste-pickers with entry cards to the local waste site, but not the ID cards they had been seeking (iTC2).

Other uses were also limited, and examples only given for Kibera: of international development agencies and NGOs using the data for improved planning of education resource flows, or for siting of water and sanitation facilities (iMK1b, Hagen 2017). These and the government usage descriptions highlight an issue raised by interviewees: that the upstream activity of data gathering and processing and visualisation is fairly quick, easy and "clean" in terms of local politics; but ensuring downstream use of the data let alone actual results is much more difficult, time-consuming, costly and "dirty" (iMK1a, iSKK, iTC1). Data intermediary organisations must identify potential users; build links and trust with them over time, seeking to persuade them of the value of new data; work alongside them during often-protracted and politicised sequences and iterations of decision making; and then not merely monitor results but have the power to act if implementation falls short of plans (iSKK, iTC1, iTC2, Hagen 2011).

D4. Structural Data Justice

There is evidence of some level of structural determinism in the case studies: of these pro-equity initiatives to still reflect structural inequalities, and for data to flow along the fault-lines of power in society.

A consistent theme across all initiatives was that operation of the full information value chain depended on whether or not the data had utility for powerful local actors; particularly those in local government. In Solo, the mini-atlases became an officially-mandated part of participatory budgeting because Mayor Widodo wanted his urban development innovations to work (iSKK, Bunnell et al 2013). By contrast, government officials in Pune were resistant to the participatory budgeting process, making it much harder for new data to be utilised (iOPOB1, Jobst & Malherbe 2017).

In relation to institutional forces, formal control of urban development decision processes rested with external agencies; particularly local government. Hence, it was their perception of the utility of that data which determined whether or not it would be used, in what

manner and to what end. Data also flowed according to the topography of trust in the city. Notwithstanding examples of government officials changing their view of the assets and issues in slum areas, they still tended to mistrust those communities (iSKK, iTC3). At best, substantial investment in advocacy and "bridge-building" activities by the data intermediaries was required to enable those officials to use the data provided; but at worst, even this did not stop officials from devaluing data about the community and being reluctant to use it.

All four initiatives were funded to a significant degree by international donors. That control of resources meant the initiatives were shaped by the agenda of those donors, which were for highly-visible, short-term results (iMK2, iTC1, Hagen 2017). This meant activity in all projects was a right-skewed distribution curve with rapid early growth but then tailing off to much lower levels that struggled to sustain (iMK2, iTC1). And it meant an upstream—downstream disconnect: donors would pay for production of data artefacts, especially physical and digital maps, but were much less interested in paying for the longer-term, less-visible activity of getting that data used to change urban decisions; in particular, not funding use of data by communities themselves (iMK1b, iOPOB2).

Conversely, the lack of resources among community members – mainly their financial poverty but also lack of knowledge and skills – shaped their engagement with the PEDIs. For many it meant non-engagement because of the opportunity cost that time invested in the project was time lost from earning a livelihood. So people: did not respond to requests for data, did not join in mapping of their community, did not engage in community use of data (Grillos 2017, Hagen 2017, Jobst & Malherbe 2017). And where they did engage, they brought an instrumental rather altruistic agenda; thus Map Kibera mappers came forward because the work was paid and even volunteer mappers were focused on the skill- and job opportunity-developments the work could offer (iMK1b, Berdou 2011, Lundine et al 2012).

Structural relations were most often understood in terms of the relation between state and citizens. For Map Kibera, the state's monopoly of legitimate violence had shaped data use. They avoided "provocative" uses of data that might antagonise state actors and lead to "vengeance" (iMK1b). For example, they did not gather or use data relating to housing demolitions and evictions. They saw this would potentially lead to protests that would be violently suppressed by the authorities and – reflecting state control over NGOs – lead to Map Kibera Trust being shut down. For Transparent Chennai, the lack of accountability of state institutions to citizens shaped data use. As an example, the local Slum Clearance Board and Water Corporation had no elected representatives, no public hearings, no open reporting policies (iTC1, iTC2). As a result, there were no incentives for them to change their practices and make use of new streams of community data.

Finally, in terms of epistemics, the strength of smart-city discourse shapes outcomes. With its emphasis on data and technology, it drives the idea among funders and other stakeholders that an upstream focus will be sufficient; that the addition of digital technologies into urban processes is enough to deliver urban development (iMK1b, Shekhar 2016). This makes it more difficult to engage funds and actors in the more difficult aspects of downstream processes and socio-political factors.

On the "input" side, then, there was evidence of structural determinism but looking at the "output" side this was not quite a story of full path dependency. No social structure is stable and, while there was substantial reproduction of the structures of power, datafication had enabled some incremental changes to the architecture of urban development, as discussed next.

Regarding utility, there are always competing agendas and interests in any urban development setting. The new data streams from the PEDIs assist those looking to make more open and/or more objective decisions[6]. Interests in openness were rarely found: it applied only and to only some degree to the mayoralty of Joko Widodo in Solo (Bunnell et al 2013). There were more interests in objectivity: an occasional presence in local government in Pune and Chennai; more strongly felt by external development agencies such as UNICEF and GOAL working with Map Kibera (iMK1b, iOPOB1, iTC2, Hagen 2017). And data itself – via its advocacy use by data intermediary organisations, local NGOs and local media and/or via concerns that urban decisions were visible to a wider world – could modify the perceived interests of decision-makers, nudging them in a few cases and temporarily towards greater concern for slum residents (iSKK, iTC1, Kovacic & Lundine 2013).

Institutional change was minor: in formal terms, the only regulatory change was the Solo mini-atlas incorporation; there were no examples of legal recognition of rights to land or even occupancy. In informal terms, there was some trust-building but less between citizens and government direct than between some community members and the data intermediary organisation; and then between the intermediary and some government officials. Structurally, too, insertion of the data intermediary into the landscape was the most notable feature – acting as that trust-building bridge, translating epistemes and logics, to some degree equalising the power inequalities between citizens and outside agencies (iMK1b, iTC3, Song 2016). But these intermediating relations were contingent: broken off or downgraded, for example, when new government officials came into office (iOPOB1, iTC3). And they were bounded. There were "no-go" areas that were deemed too politically-sensitive, for example if touching on corruption (iTC1, Swain 2012). They could have been the basis for holding state and other actors to account except, as noted, data on actions and results – on the extent to which new infrastructure, services and resources were being delivered into the slums – was not being gathered (iTC2, Menon 2013, Feruglio & Rifai 2017).

Epistemic change was also bounded. Communities to some extent, and external actors to a greater extent, had better knowledge of these informal settlements. We noted evidence of "slum imaginaries" of some of those external actors changing to a more positive view. But negative images persisted, at least among some local government officials (iTC3, Menon 2013).

Lastly, the key resource change is the availability of data. There were barriers to this flowing to slum residents; having to be converted to a less-valuable analogue (i.e. paper-based) form due to lack of access to digital technologies within slums. Meanwhile, being open and

---

[6] Data is never completely open or neutral or objective, but it is the perception that PEDI data is more open and more objective than previous data that matters here.

digital, it flowed more readily to others outside the community. And as a usable resource, it favoured those with the other resources – skills, knowledge, time, confidence, money – necessary to make the value chain operate; resources which the marginalised community lacked. The impact of that use could benefit the community: the data intermediaries empowered to better advocacy on behalf of communities; the international NGOs empowered to planning better interventions. But control and empowerment still resided outside the community. In particular, it is the data intermediary organisations that are empowered – interrogating openness, one finds that it is often sub-sets or summaries or visualisations of data that are being circulated. Control over the foundational dataset including ownership and rights to update typically rests with the data intermediaries.

E. Discussion and Conclusions

E1. Distributive Data Justice

Distributive data justice – the concern for who gets what as a result of data systems – is a lens applied across the other dimensions (see Figure 1). It has particular relevance here given the foundational concern about marginalisation, and the injustices of data-based inequality suffered by the slum communities: their location, assets, issues are less visible than those of more affluent areas and residents, with a knock-on to inequality of infrastructure, service provision and political role.

Procedurally, the overall picture could be seen as a pyramid. At the base, the majority of community members are not involved in the data-information-decision-action processes of the information value chain. Above them, the largest number of beneficiaries are within the community but gain only the interpretive value of data; the mappers who capture data are far fewer but gain more; the midstream professionals are fewer still but gain greater capabilities; and the intermediary organisations benefit most of all.

A similar picture can be seen in terms of visibility. The most marginalised – those without homes, those without identification, those residing on the physical margins who come in to the city to make their living – are rarely made visible by these data initiatives. Conversely, data that might truly challenge political elites e.g. enabling them to be held to account, remained largely invisible. The communities which are made visible reap some benefit directly but they lose control of their representation; becoming legible to others who can make use of the community's data twin for their own purposes. From this, we can see the ambivalence of legibility. Slums must be legible to government, NGOs, development agencies, etc if they are to benefit from the resources, services, support, etc that these external organisations offer. But that same legibility exposes slum residents to any other agendas these organisations may hold; agendas that may be orthogonal to slum interests or even counter to those interests: extractive, persecutory or predatory. Even if direct evidence of the latter was limited, the fear of it led some data intermediaries to avoid gathering certain types of data, and led some groups to prioritise their right to privacy and either resist or refuse to engage with data gathering.

External agendas have served the slums: these data initiatives bring better-planned and more resources and services. But alongside the glass half-full is a glass half-empty. These absolute improvements are always less than intended or needed, and sometimes merely symbolic or temporary. Marginalised groups find some greater voice and place in urban decision making than previously thanks to these data initiatives. But, particularly in city-wide initiatives as in Pune and Solo, relative inequality still grows as more formalised areas and wealthier residents retain a stronger voice and benefit more.

This overall picture is both explained by and reflected in the structural data justice evidence. Path dependency is the main narrative: the wider structural inequalities of power-interest significantly shape the data systems and operations of the information value chain. In turn, the value chain largely reproduces that wider configuration and constrains pro-equity aspirations. This is not a completely static picture: the value chain does reshape – subtly, incrementally – the structural determinants of power and utility. There is an incremental empowerment in terms of new data access, external perceptions of self-interest and of slums and their residents, and the landscape of organisations and trust. But there is no evidence of a wider transformation of the substrates of urban inequality, with communities potentially more dependent on external forces as a result of their heightened legibility.

E2. Conclusions and Recommendations

Datafication is a key trend within international development; with effects seen in every development sector and effects growing every day. Yet analysis of real-world experiences and their breadth of impact has to date been limited; in part due to a lack of analytical frameworks.

This paper therefore makes two significant contributions. First, it exposes a full picture of the impact of one type of pro-equity data initiative – community mapping – on those who are marginalised within the city. New data flows do have an impact but disproportionately serve those with the motivation and power to use that data. Results are certainly beneficial for slum communities and other marginalised citizens, and these initiatives can be justified on that basis. However, though there can be no exact calibration from qualitative research, it is likely that these pro-equity initiatives actually increase relative inequalities. Ordinary community members have seen some benefits but external actors who find the data to match their agenda and capabilities, benefit more. It is the latter who are more empowered to access, use and control the new data.

Second, and in order to perform this evaluation, the paper presents and demonstrates an explicit, systematic and comprehensive framework for the analysis of data systems and initiatives. The framework is data-based: acknowledging the important growth in both reality and discourse around datafication by placing data at the heart of the model. But it simultaneously decentres data and digital technology by guiding attention to wider processes, impacts and structures. There was some overlap in the dimensions used; with the rights discussion particularly touching on procedural, instrumental and structural issues. However, in general the perspectives were additive; offering new insights and particularly substantiating the need to incorporate and understand context – rights, structures, interests – in order to fully understand the implications of datafication.

The framework was applied here to analyse just one type of data initiative. It should be equally applicable to all types of datafication – urban and otherwise – but that application must form part of a future research agenda. Another part will be action research; working with data intermediary and other organisations in using the framework to guide design and implementation of data initiatives. In advance of that, one can – building from Fox (2015) – identify three approaches to pro-equity data initiatives: technical, social and critical.

A technical approach focuses just on the digital data and technology, and the findings here show this to be too limited a view that falls short both in terms of conceptualisation and impact. A social approach focuses on achievement of development results. Findings here suggest such an approach should recognise the procedural benefits that can emerge alongside instrumental results. And they suggest that such projects need a clear theory of change, with the information value chain providing the basis for that. Project design can either look for or seek to create a functioning information value chain; one that has utility as its driver and the necessary resources present as enablers.

A critical approach would encompass the wider context of rights, power and interests. The findings suggest value in modesty of expectations for what can be achieved through pro-equity data initiatives; accepting incremental rather than transformational outcomes; accepting absolute improvements in informal settlements, even if broader inequalities may widen. They highlight the importance of identifying tensions and disjunctures within the context: seeking out NGOs that see an advantage over others within the competitive funding environment if they make use of new or more-objective data; looking for newly-appointed mayors or urban commissioners or development agency directors who bring a new agenda that new or more-objective data could be seen to assist. And they highlight the important structural role performed by data intermediary organisations: as channels through which citizens can influence the balance of representation/privacy and the nature of data ownership and access; as important determinants of the constitution and distributive benefits of information value chains.

More generally, the findings support the value of imbuing data projects with the values of data justice; ensuring their design and implementation follow principles of procedural, rights-based, instrumental, structural and distributive justice. Such principles may be found, for example, in the "Data-Justice-for-Development Manifesto" (see Box 1; Heeks 2017b). These can be a guiding framework, particularly for data intermediary organisations given they are focal nodes for emancipatory action.

Box 1. A Data-Justice-for-Development Manifesto

1.   Demand just and legal uses of development data.

2.   Demand data consent of citizens that is truly informed.

3.   Build upstream and downstream data-related capabilities among those who lack them in developing countries.

4.   Promote rights of data access, data privacy, data ownership and data representation.

5.   Promote data system outcomes that address international development goals and priorities; including the goals and priorities of data subjects.

6.   Support "small data" uses by individuals and communities in developing countries.

7.   Advocate sustainable use of data and data systems.

8.   Create a social movement for the "data subalterns" of the global South.

9.   Stimulate an alternative discourse around data-intensive development that places issues of justice at its heart.

10. Develop new organisational forms such as data-intensive development cooperatives.

11. Lobby for new data justice-based laws and policies in developing countries (including action on data monopolies).

12. Open up, challenge and provide alternatives to the data-related technical structures (code, algorithms, standards, etc) that increasingly control international development.

References

Baud, I. (2016). Digitisation and participation in urban governance. In *Local Governance, Economic Development and Institutions*, G.M. Gomez & P. Knorringa (eds), Palgrave Macmillan, Basingstoke, UK, 86-97.

Benequista, N. (2015). Journalism from the 'Silicon Savannah'. *Stability*, 4(1), 1-16.

Berdou, E. (2011) *Mediating Voices and Communicating Realities*. IDS, Brighton, UK.

Bunnell, T., Miller, M.A., Phelps, N.A. & Taylor, J. (2013). Urban development in a decentralized Indonesia. *Pacific Affairs*, 86(4), 857-876.

Chakraborty, A., Wilson, B., Sarraf, S. & Jana, A. (2015). Open data for informal settlements. *Journal of Urban Management*, 4(2), 74-91.

Dalton, C. & Thatcher, J. (2014). What does a critical data studies look like, and why do we care? *Society + Space*, 29.

Donovan, K. (2012). Seeing like a slum. *Georgetown Journal of International Affairs*, Winter/Spring, 97-104.

Feruglio, F. & Rifai, A. (2017). *Participatory Budgeting in Indonesia*. Making All Voices Count.

Fox, J.A. (2015). Social accountability. *World Development*, 72, 346-361.

Grillos, T. (2017). Participatory budgeting and the poor. *World Development*, 96, 343-358.

Hagen, E. (2011). Mapping change. *Innovations*, 6(1), 69-94.

Hagen, E. (2017). *Open Mapping from the Ground Up*. Making All Voices Count.

Haklay, M. (2013). Neogeography and the delusion of democratisation. *Environment and Planning A*, 45(1), 55-69.

Heeks, R. (2017a). Data justice, presentation at IFIP WG9.4 conference *ICTs for Promoting Social Harmony*, Yogyakarta, Indonesia, 22-24 May.

Heeks, R. (2017b). *A Structural Model and Manifesto for Data Justice for International Development*, GDI Development Informatics Working Paper no.69. University of Manchester, UK.

Heeks, R. (2018). *Information and Communication Technology for Development*. Routledge, Abingdon, UK.

Heeks, R. & Renken, J. (2018). Data justice for development. *Information Development*, 34(1), 90-102.

Jobst, M. & Malherbe, M. (2017). *Participatory Budgeting in Pune*. Uppsala University, Sweden.

Kitchin, R. (2014). The real-time city? *GeoJournal*, 79(1), 1-14.

Kitchin, R. & Lauriault, T. (2018). Towards critical data studies. In *Thinking Big Data in Geography*, J. Thatcher, J. Eckert & A. Shears (eds), University of Nebraska Press, Lincoln, NE, 3-20.

Kovacic, P. & Lundine, J. (2013). Mapping Kibera. In *Bits and Atoms*, S. Livingston & G. Walter-Drop (eds), Oxford University Press, Oxford, 115-129.

Krishnamurthy, R. & Desouza, K.C. (2015). Chennai, India. *Cities*, 42, 118-129.

Lokanathan, S. (2017) *Mapping Big Data for Development and the Global Goals*. LIRNEasia, Colombo.

Lokanathan, S., Kreindler, G.E., de Silva, N.N., Miyauchi, Y., Dhananjaya, D. & Samarajiva, R. (2016). The potential of mobile network big data as a tool in Colombo's transportation and urban planning. *Information Technologies & International Development*, 12(2), 63-73.

Lundine, J., Kovacic, P. & Poggiali, L. (2012). Youth and digital mapping in urban informal settlements. *Children Youth and Environments*, 22(2), 214-233.

Menon, S. (2013). *Participatory Budgeting in Pune*. Centre for Environment Education, Pune.

Neff, G., Tanweer, A., Fiore-Gartland, B. & Osburn, L. (2017). Critique and contribute. *Big Data*, 5(2), 85-97.

Obermayr, C. (2017). *Sustainable City Management.* Springer, Cham.

Otiso, K.M. (2002). Forced evictions in Kenyan cities. *Singapore Journal of Tropical Geography*, 23(3), 252-267.

Padmanabhan, V., Raman, N.V. & Shekhar, S. (2013). *Plan B: Practical Strategies to Improve Planning in Indian Cities*. Indian Council for Research on International Economic Relations, New Delhi.

Panek, J. & Sobotova, L. (2015). Community mapping in urban informal settlements. *Electronic Journal of Information Systems in Developing Countries*, 68(1), 1-13.

Patel, S. & Baptist, C. (2012). Documenting by the undocumented. *Environment and Urbanization*, 24(1), 3-12.

Pfeffer, K. & Verrest, H. (2016). Perspectives on the role of geo-technologies for addressing contemporary urban issues. *European Journal of Development Research*, 28(2), 154-166.

Rifai, A., Asterina, N. & Hidayani, R. (2016). *Improving the Transparency, Inclusivity and Impact of Participatory Budgeting in Indonesian Cities*. Making All Voices Count.

Samarajiva, R., Lokanathan, S., Madhawa, K., Kreindler, G. & Maldeniya, D. (2015). Big data to improve urban planning. *Economic & Political Weekly*, 50(22), 43-48.

Sanchez, C., Mackaness, W. & McLaren, R. (2013). Sustainable mapping projects, paper presented at *GISRUK 2013*, Liverpool, 3-5 Apr.

Scott, J.C. (1998). *Seeing like a State*. Yale University Press, New Haven, CT.

Sengupta, R., Heeks, R., Chattapadhyay, S. & Foster, C. (2017). *Exploring Big Data for Development*, GDI Development Informatics Working Paper no.66. University of Manchester, UK.

Shekhar, S. (2016). Introduction: (re)prioritising citizenship in smart cities governance. In *(Re)Prioritizing Citizenship: Setting a New Agenda for Smart Cities Governance*, S. Shekhar (ed), CAG, Chennai, 5-8.

Shekhar, S. & Padmanabhan, V. (2015). *The Quality of Civic Data in India And the Implications on the Push for Open Data*. Transparent Chennai, Institute for Financial Management and Research, Chennai, India.

Shkabatur, J. (2014). Interactive community mapping. In *Closing the Feedback Loop*, S. Bailur & B. Gigler (eds), World Bank, Washington, DC, 71-106.

SKK (2010). *Gilingan Neighbourhood Map*. Solo Kota Kita, Solo.

Song, L.K. (2016). Planning with urban informality. *International Development Planning Review*, 38(4), 359-381.

Spratt, S. & Baker, J. (2015). *Big Data and International Development*. IDS, University of Sussex, UK.

Swain, S.S. (2012). The unequal access to municipal services and the role of local elected representatives, paper presented at *N-AERIS XIII*, Paris, 22-24 Nov.

Taylor, L. (2017). What is data justice? *Big Data & Society*, 4(2).

Taylor, L. & Broeders, D. (2015). In the name of development. *Geoforum*, 64, 229-237.

Taylor, L. & Richter, C. (2015). Big data and urban governance. In *Geographies of Urban Governance*, J. Gupta, K. Pfeffer, H. Verrest & M. Ros-Tonen (eds), Springer, Cham, 175-191.

Townsend, A.M. (2013). *Smart Cities*. W.W. Norton, New York, NY.

UN-Habitat (2016). *World Cities Report 2016*. UN-Habitat, Nairobi, Kenya.

UN-Habitat (2017). *New Urban Agenda*. UN-Habitat, Nairobi, Kenya.

van Veenstra, A.F., Esmeijer, J., Bakker, T. & Kotterink, B. (2014). *Data and the City*. TNO, Delft.

Willis, K. (ed) (2017). *Whose Right to the Smart City?* Plymouth University, UK.