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

Paper No. 69

A Structural Model and Manifesto for Data Justice for International Development

RICHARD HEEKS

2017

ISBN: 978-1-905469-71-0

Published by: ***Centre for Development Informatics***
Global Development Institute, SEED
University of Manchester, Arthur Lewis Building, Manchester, M13 9PL, UK
Email: cdi@manchester.ac.uk Web: http://www.cdi.manchester.ac.uk

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

A Structural Model and Manifesto for Data Justice for International Development

Richard Heeks
Centre for Development Informatics, University of Manchester, UK

2017

Abstract

With growing use of data in international development, there is growing interest in data justice. One argument is that this must best be understood in terms of structural data justice (SDJ): the degree to which society contains and supports the data-related institutions, relations and knowledge systems necessary for realisation of the values comprised in a good life. But only hypothetical models of SDJ have been proposed to date.

The purpose of this paper is to take one of the proposed SDJ models and revise it on the basis of experience with field studies of big data and other new data streams in India and Kenya. Those field studies produced three tests of a data justice model, asking whether it can encompass: the impact of social structure on data systems; the impact of data systems on social structure; and the role of datafication and related technological affordances. On the basis of the three tests, a revised and improved model of structural data justice is developed, which is commended as a conceptual frame to use in future research on data-intensive development.

The model is shown to incorporate all types of data justice, and to be of particular value to critical data studies in understanding how both "power over" and "power to" are exercised in data-intensive development. The model is also the basis for derivation of a "Data-Justice-for-Development Manifesto", which can be used to guide development policy and practice.

A. Introduction

We are moving towards an era of "data-intensive development": "the growing presence and application of data in the processes of international development" (Heeks 2017a). For example there is growth in availability and use of big data (Hilbert 2016), and there is growth in availability and use of open data (WWW 2017). With this expansion of data-intensive development have come a number of development benefits: faster and better decisions that are facilitating improved development outcomes in health, agriculture, urban planning, etc (Cartesian 2014, Kshetri 2014). But there have also been a number of growing concerns about emerging negative impacts: loss of privacy, discrimination, growth in inequality, etc (Spratt & Baker 2015, Taylor & Broeders 2015).

Those concerns have been framed in a number of ways but one strand of work has linked itself to "data justice": "the specification and pursuit of ethical standards for data-related resources, processes and structures" (Heeks 2017b). This is still a very formative topic area with a very few papers emerging in recent years (e.g. Newman 2015, Dencik et al 2016) including one paper specifically addressing data justice in the context of international development (Heeks & Renken 2017).

This latter paper had a specific argument. Having provided evidence of data injustices occurring in developing countries, it identified three mainstream perspectives on the converse, data justice (*ibid.*:3):

- "Instrumental data justice means fair use of data; it therefore focuses on the outcome of use of data"
- "Procedural data justice means fair handling of data" along all parts of the information value chain (see Figure 1; Heeks 2017c).
- Distributive rights-based justice encompasses rights of privacy, access, ownership and representation; the enactment of which shapes distribution of data resources.

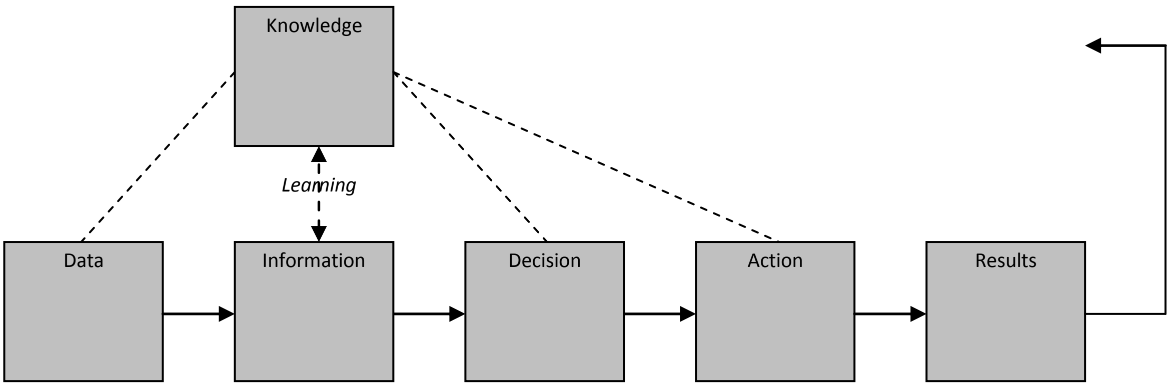

Figure 1. The information value chain

However, the earlier paper argued that these mainstream views of data justice had a number of shortcomings. In particular, a failure to "encompass the social structures which at least partly determine data uses, processes, distributions, and rights" (Heeks & Renken 2017:7). It further argued that these structures needed to be encompassed and that "the foundation of data justice must be structural data justice, which we can define as 'the degree to which society contains and supports the data-related institutions, relations and knowledge systems necessary for realisation of the values comprised in a good life'" (*ibid.*:7).

The paper then went on to propose three different ways in which this idea of structural data justice could be conceptualised:

- A connectivist view, building on "the work of Iris Marion Young who takes a network view of social structure – what she calls her 'social connection model' – seeing individuals occupying particular social positions" (*ibid.*:8).
- A critical view looking at the competing interests that intersect around data systems and which have different relative degrees of power with which to shape the design and outcome of those systems.
- A capabilities view building from the work of Amartya Sen.

Readers are referred to the Heeks & Renken paper for further detail. However, that earlier paper was developed solely on the basis of secondary data, and its models for structural data justice were just propositions for future testing. This current paper reports on subsequent developments; in particular, a set of field studies conducted during 2016 and 2017. The purpose of this paper is to take one of the earlier-proposed models of structural data justice and to review and develop it in light of the field studies.

The paper proceeds as follows. First, a short evidence section will outline the field studies undertaken. Then the main body of the paper reviews a basic model of structural data justice and steadily refines it on the basis of field study evidence, which is presented as a set of three tests for a framework of structural data justice. The paper then reflects on further elements in the Heeks & Renken paper: the extent to which the revised model captures other proffered conceptualisations, and finally the implications of the revised model for practice; specifically in deriving a proposed "Data-Justice-for-Development Manifesto".

B. Field Study Evidence

The evidence reported below derives from four field studies of data-intensive development. Three were coordinated by Sumandro Chattapadhyay of the Centre for Internet and Society, Bengaluru, India:

- A study of big data in an anonymised state electricity corporation, "Stelcorp" based in "Janakari" state. The study was led by Ritam Sengupta and is published as Sengupta et al (2017). Losses in electricity distribution – both technical losses plus non-payment of bills – led Stelcorp to install digital meters (some online, some offline that had to be read via human intervention) for all consumers and electricity transformers. The terabytes of data generated are used to some extent in billing, rectification of faults, and planning improvements to the electricity network.
- A study of big data in the Bengaluru Metropolitan Transport Corporation (BMTC), which runs the city's 6,000-plus buses. The study was led by Vanya Rakesh and is available as Rakesh et al (2017). The Corporation has financed development of a new big data system, called the Intelligent Transport System (ITS). The main elements of ITS are

online, real-time tracking of all buses, and electronic ticketing machines from which all ticketing data is uploaded. The gigabytes of data generated are used to feed a passenger information system that is intended to show where buses are and when they will arrive at bus stops; and to support operational management such as cash management of ticketing. It is the future intention within BMTC that data should be used for tactical management such as rationalisation of bus routes and schedules.
- A study of big data in the school system in Andhra Pradesh state, India, led by Anuj Srivas. The big data project is still in progress because it involves gathering of performance and other data on tens of thousands of secondary school students over multiple years. The intention is to use machine learning to then predict which students are likely to drop out of school, and to intervene with those identified as being at risk.

The fourth study was coordinated and led by Satyarupa Shekhar of the Citizen Consumer and Civic Action Group, Chennai, India. Not yet published at the time of writing, it reports the experiences of two initiatives that developed data relating to marginalised urban communities. Map Kibera facilitated the digital mapping of Kibera slum in Nairobi by community members. Transparent Chennai was a basket of projects that gathered data on various issues in the city: homelessness, availability of public services, vacant land available for development, and services in the tenement blocks provided for evicted slum dwellers.

C. Developing a Structural Model of Data Justice

As explained above, the Heeks & Renken paper proposed a number of possible ways forward for conceptualising structural data justice. The foundation model used here will be the basic one proposed from critical data studies, reproduced in Figure 2.

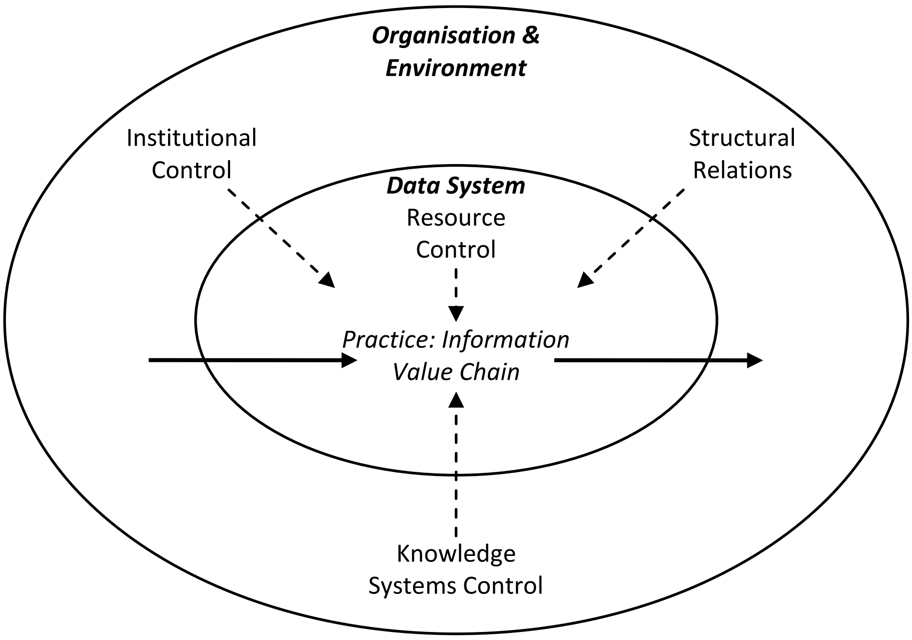

Figure 2. Foundation model of structural data justice

In this section, we review this model by considering evidence from the field studies. Derived from the studies are three tests for any model of structural data justice. These are used to develop a more comprehensive – also of course more complex – model of structural data justice.

Test 1. Can the model encompass the impact of pre-existing structure?

It is, of course, a core argument of structural data justice that structural features shape other aspects of data justice. Many different definitions of social structure are possible but the approach taken here sees social structure in terms of power; particularly in terms of "power over": the ability of powerful actors – because of the disposition of social structure – to shape other aspects of data justice. Drawing from the literature on social structure and power (e.g. Hardy & Phillips 1998, Hearn 2012), power is seen to come from control over resources and practices but also from three other things that themselves determine control over resources and practices: control over institutions, position within structural relations, and epistemic control over knowledge systems. It is this view which led to the definition of structure above as "data-related institutions, relations and knowledge systems" given that resources and practice were already encompassed in distributive and procedural notions of data justice.

We can see examples in practice of this structural shaping of data justice. In the "Janakari" state electricity case, installation of digital electricity metering for consumers could have led to a sharp reduction in financial losses for the electricity corporation. This would have occurred as accurate data on electricity usage led to accurate billing for all consumers, and thence to more effective collection of amounts owed. This did happen in urban areas but not elsewhere. Instead, there was estimated to have been a growth in those not paying or underpaying their electricity bills in rural areas with potential for inequality given "it was the richer and better-politically-connected sections in Janakari rural areas – rural industrialists, larger-scale farmers, larger-scale irrigation owners – who were most likely to be appropriating free or low-cost electricity" (Sengupta et al 2017:20). The explanation is that this big data initiative was inserted into the heavily-politicised nature of electrification in India. Those with political power – politicians with power by virtue of their institutional control and structural position – saw that big data was increasing billing compliance rates in urban areas, and was also leading to fewer blackouts and brownouts. This enabled the politicians to drive a large-scale programme of electrification in rural areas, with an implicit notion that higher urban revenues could cross-subsidise not just rural expansion but also rural non-payment. This was a programme that they knew would win votes, especially when more powerful consumers were allowed to circumvent the accurate billing that new metered data streams could provide.

In the electricity case, wider structures impacted the way big data was being used (thus linking to procedural data justice) and to the outcomes of that data use (thus linking to instrumental data justice). The field studies also exposed instances of wider structures shaping distributive rights-based justice: privacy, access, ownership and representation. For example, the three sectoral case studies undertaken in India – electricity, transport and education – involved public sector organisations. They gathered data from citizens – electricity consumers, bus users, school students – and they gathered data about public services. Yet in none of those cases to date has the data been made available in some form of open data. This has arisen because such data access goes against the institutional norms of the organisations involved, and more directly because it goes against the interests of the powerful figures within those organisations, who fear the impact of data-driven transparency.

From these examples and others we could see ways in which the four elements of structure (structural position, and resource, institutional and epistemic control) influence the practices of the information value chain; thus validated the model above. However, we can also see some shortcomings of the model. The influence of structure on procedural data justice is well-reflected. But the influence on instrumental data justice – the outcomes of use of data such as those in the Indian electricity case – is not. And nor is the influence on distributive rights-based justice such as the right to access. Development results therefore need to be acknowledged as an output from the information value chain. And data rights need to explicitly appear.

Second, structure has to be enacted. In the cases noted above, it was actors with structural power who shaped the data system and its practices. From the structural elements included in the model we can understand how they are able to shape data rights, practices and outcomes through their structural power. But the cases were driven by the particular interests of these actors: the desire to win votes, or the desire to avoid scrutiny that explain why they seek to shape data rights, practices and outcomes. This driving force shapes the information value chains that do run and those that do not, and it could be given various names such as interests, incentives or values. Here, I suggest using the term "utility" to represent the combination of extrinsic incentives and intrinsic motivations and beliefs that shapes the goals that actors value and seek to achieve through the information value chain.

These developments are incorporated into the model, as shown in Figure 3.

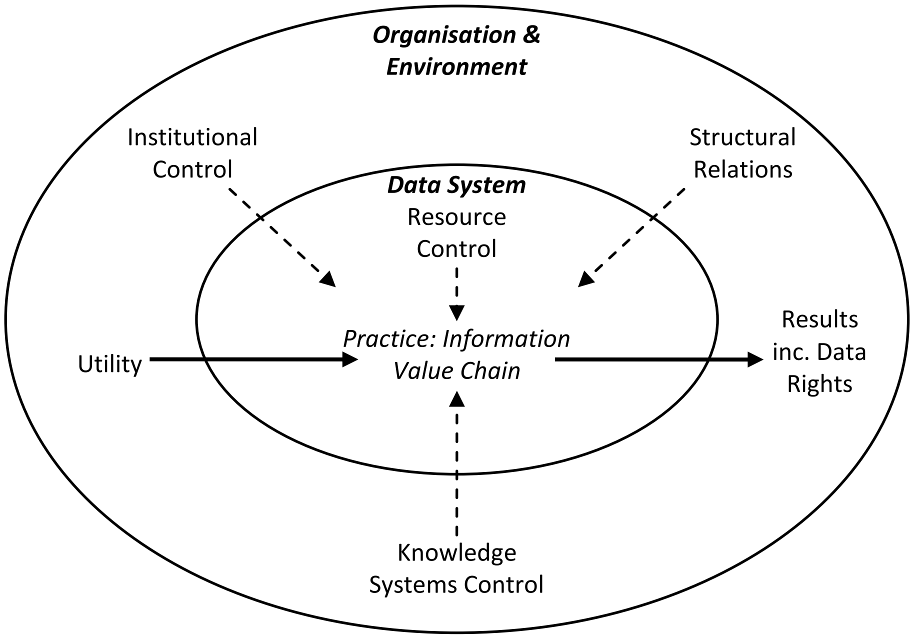

Figure 3. First iterative development of structural data justice model

Test 2. Can the model encompass the impact of data systems on structure?

The picture painted so far has been rather static and uni-directional: focusing on ways in which the justice or injustice of wider structures impacts other aspects of data justice. But structure is not static. It is influenced by data systems and the practices of the information value chain.

For example, the advent of big data in the Indian electricity and transport cases has been associated with a change in structural relations within the respective organisations, and an exacerbation of pre-existing structural inequalities. In both cases, there have been two upward shifts of power towards central management: from labour to management, and from middle to central management. In the electricity corporation, for instance, introduction of online metering has already led to 40% of meter-readers losing their jobs and those that remain capture a declining proportion of the corporation's overall data. In the bus transport corporation, "Big data systems enable central management to directly access performance data from the front-line of operations, and to automatically undertake and communicate performance management; obviating the need for intermediating management layers" (Rakesh et al 2017:18). As a result the corporation's divisional offices – the layer that sits between central management and the bus depots – are being closed: a significant organisational restructuring.

To take another example, we have seen the way in which all the data initiatives in India have changed the surrounding epistemics. Data systems do this because they create a separate virtual model of the phenomena about which they gather data: the so-called "data double" or, for place-based representations, the "shadow map" (Taylor & Broeders 2015). This in turn can alter the "imaginaries" of those involved – the mental model and worldviews they have about the phenomena – and the wider discourse about those phenomena. For example, big data in the bus transport corporation has changed the mental model of managers: "the daily operations of the bus fleet and bus crews were largely opaque to management prior to ITS, but they are becoming increasingly visible and thus changing the perceived picture of BMTC that managers hold in their heads" (Rakesh et al 2017:17).

In the urban data initiatives, we also found examples of data-enabled epistemic reframing. In the case of Transparent Chennai, the data gathered from informal settlements enabled some reorientation of perceptions and discourse among urban planners and officials. They moved somewhat from an epistemic of the troublesome slum-dweller that needed eviction, to a discourse about informal residents who had rights including a right of access to public services.

Of course by what they make visible and spotlight, data systems may throw into further shadow that which is not visible through digital data. In the bus transportation case, datafication of the buses and ticketing has reinforced a discourse that focuses on the (negative) actions of the bus crews: speeding, missing bus stops, diverging from the set route by bus drivers; failure to issue tickets and theft by the bus conductors. Framing these as the key problems of bus transportation takes the discourse away from other issues such as complaints about corruption within operational and strategic management of the corporation. Similarly in one urban data case, data was gathered on the quality of tenements to which evicted slum dwellers had been relocated. By strengthening the discourse around improvements needed in the tenements, this initiative rendered conditions in informal slums and the process of eviction less visible, and implicitly legitimised eviction as an urban development strategy in Chennai.

The impact of data systems on wider structures is acknowledged in the model through the arrow into the outer domain. However, it would seem better to provide a more explicit recognition by changing the links between structure and data practice to two-way interactions. In focusing on the two-way relation between data systems and social structure, the position of data rights comes under further analysis. It is not merely that rights are impacted by data systems; they also shape those systems. Indeed, data rights are part institution, part episteme: for instance, the norms and discourse around rights of access to public sector data are part of what is shaping the pressure on various Indian organisations to open up their datasets (albeit, they have resisted so far for the reasons noted above). Finally, given the focus on the language of "epistemic", it would seem simpler to re-phrase "knowledge systems control" to "epistemic control".

We can therefore move to a second iterative development of the model, as shown in Figure 4.

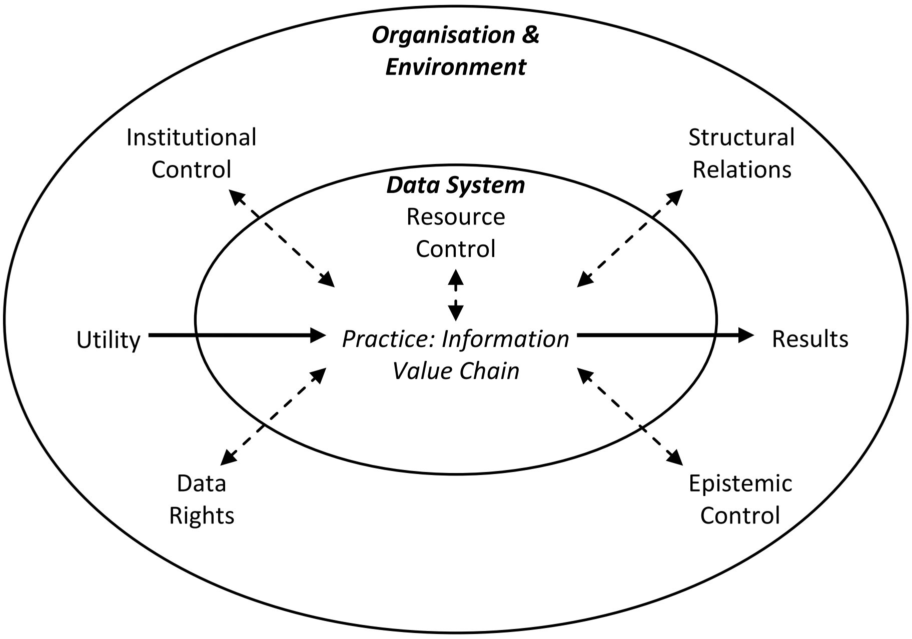

Figure 4. Second iterative development of structural data justice model

Test 3. Can the model encompass datafication?

In the examples given above, the technology and its capabilities play a central role. It is the digitising capabilities of the online meters that lead to the declining requirement for human meter-readers. It is the datafying capabilities of the Intelligent Transport System that create the data double of the bus system that enables a new imaginary in the minds of managers. Yet these capabilities are not well-reflected in the model so far.

To understand these, we can differentiate capabilities into the functionalities and the affordances of data systems. The core functionalities of new data systems can be described in terms of the much-used 3V data qualities: that they provide a greater volume, velocity and variety of data than previous systems. "The qualities are inherent functionalities of data. From these qualities, combined with purposive use by individuals or organisations, the following affordances emerge [*developed from Lycett (2013) and Nambisan (2016)*]:

- ***Datafication***: an expansion of the phenomena about which data are held. A greater breadth: holding data about more things. A greater depth: holding more data about things. And a greater granularity: holding more detailed data about things. This is accelerated by the second affordance . . .
- ***Digitisation***: not just the conversion of analogue to digital data but the same conversion for all parts of the information value chain. Data processing and visualisation for development becomes digital; through growth of algorithms, development decision-making becomes digital; through growth of automation and smart technology, development action becomes digital. Digitisation means dematerialisation of data (its separation from physical media) and liquification of data (its consequent fluidity of movement across media and networks), which underlie the third affordance . . .
- ***Generativity***: the use of data in ways not planned at the origination of the data. In particular, data's reprogrammability (i.e. using data gathered for one purpose for a different purpose); and data's recombinability (i.e. mashing up different sets of data to get additional, unplanned value from their intersection)" (Heeks 2017a).

Without consideration of these affordances, the model of structural data justice would give too little weight to the technology that lies at the heart of both structure and process; would fail to acknowledge the "data-ness" of structural data justice. To understand how to incorporate this, we can use the definition of affordances as "the potential actions an individual or organisation with a purpose can undertake with the [*data*] system within the context of the environment within which they function" (Heeks 2017c). With purpose already included via utility, affordances therefore act as a filter between utility and practice; delimiting but also enabling the practices that can be undertaken. For the sake of simplicity, in Figure 5, I will represent in terms of the first affordance: datafication; but see this as standing for all the affordances of data systems.

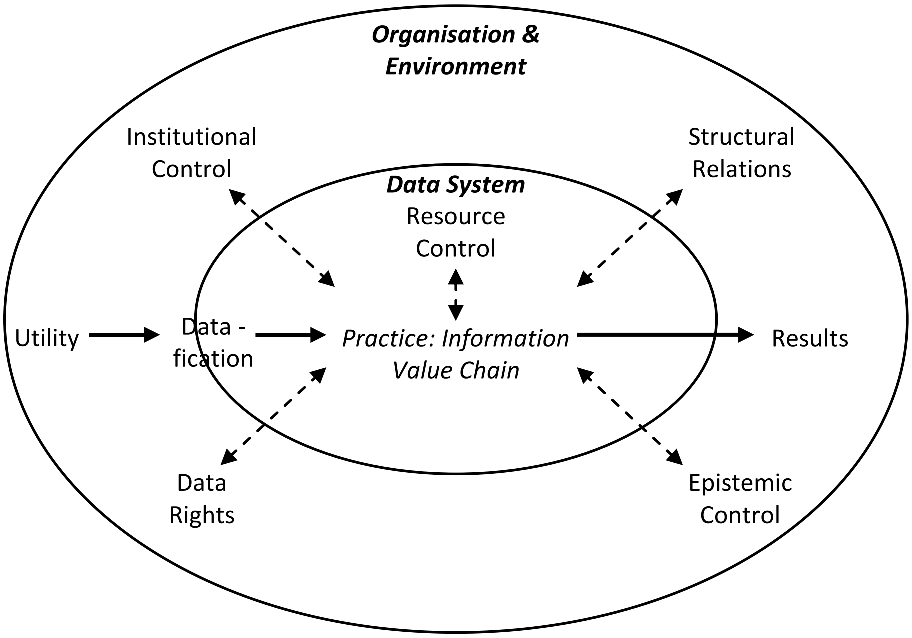

Figure 5. Revised structural data justice model

D. Discussion

The revised structural data justice model therefore represents the recommended starting point for future research on data and development. It can be used pre hoc to design research, and as a post hoc analytical lens to re-analyse existing cases of data-intensive development. Nonetheless, we can still interrogate the model further.

Having developed the model of structural data justice on the basis of field experience, we can check where the four forms of data justice fit into the model. In doing this – see Figure 6 – it can be seen that all four forms are represented, but also that the separation of types of justice may be somewhat artificial. For example, data rights are part of broader social structure, and impacts of data systems at least partly affect broader social structure.

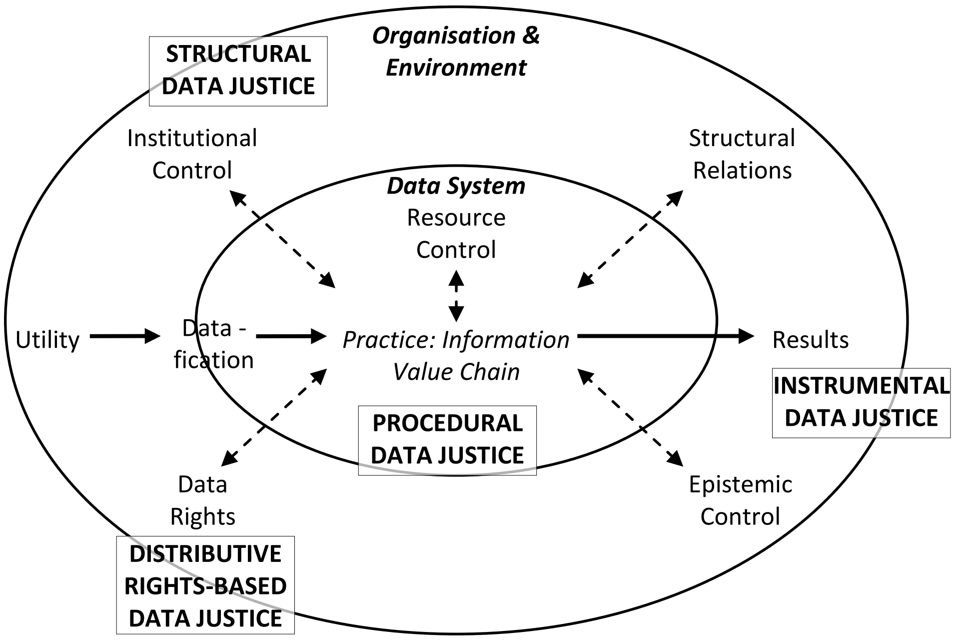

Figure 6. Forms of data justice within revised SDJ model

What about the other conceptual routes proposed in the Heeks & Renken paper: how does this model relate to them? In terms of Young's social connection model then the revised SDJ model (Figure 5) clearly encompasses that in its consideration of structural relations. However, it considers this network / positional perspective as only one aspect of social structure so is hard to argue as a direct operationalisation of Young's model.

Such an argument might be easier to sustain in the case of critical data studies. Indeed, the origins of the foundational (Figure 2) model were as an illustration of critical data studies. A useful concept here is the idea of "data assemblages": "all of the technological, political, social and economic apparatuses that frame [*the*] nature, operation and work" of data systems (Kitchin and Lauriault, 2014:6). One could readily make the case that institutions, structural relations, knowledge systems and rights represent those assemblages. The SDJ model does not separately identify technology as part of the assemblage but technological infrastructure can be recognised as both an institutional and structural element, which provides the datafication and related affordances. Development of the foundational model to include utility does also acknowledge the notion of actor interests and the way in which those are either strengthened or weakened through the operation of data systems; an important focus for critical data studies.

Where the model is less clear is in the idea of conflicting interests: there is no obvious dialectic and yet those do emerge from the field studies in relation to the views and interests of differing groups. In implementation of the model, one would therefore need to ensure a critical methodology such as critical systems heuristics which asks: who gets what; who controls what; who does what; who gets affected by the process, and with what justification (adapted from Reynolds 2014).

D1. Data-Related Agency and Capabilities

As noted in Heeks & Renken (2017), a danger of structural approaches is structural determinism that misses the role of human agency. The incorporation of practice and the presence of decisions and actions within the information value chain mean the revised SDJ model can analyse human action. But the essence of agency is a freedom and autonomy to act. As an input to the model, this “power to” does derive from control over institutions, resources, episteme, etc, as shown in the SDJ model but it is not specifically identified. And in relation to the impact on agency this is just subsumed within “results”: both the agency of those who are directly using the informational or decisional outputs of the data systems (such as electricity corporation managers planning extensions to the distribution network), and those indirectly affected by the emergent decisions and actions (such as tenement residents whose service access is improved).

To bring this into the model, Sen’s ideas of capabilities and functionings can be included, as shown in Figure 7.

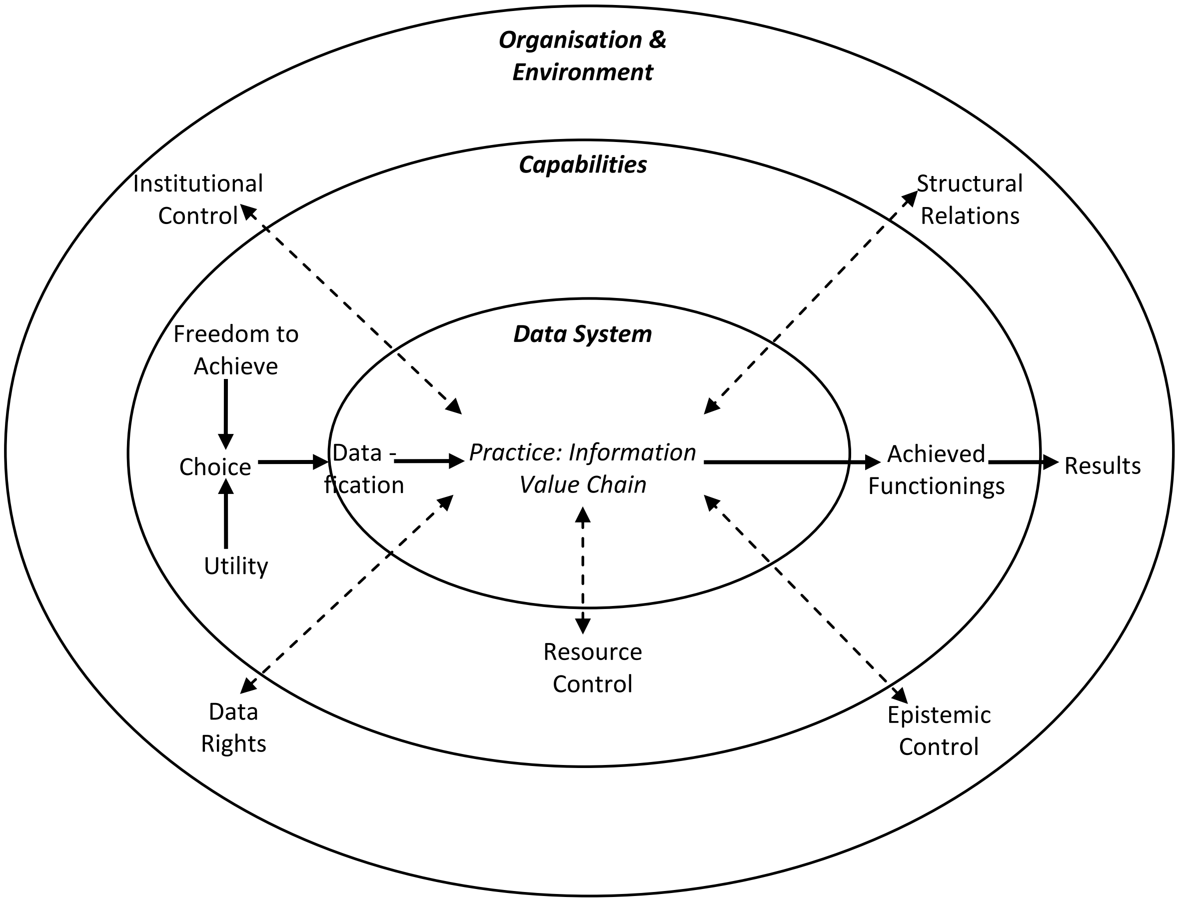

Figure 7. Incorporating explicit agency and capabilities into the SDJ model

The model remains significantly structural – appropriately so given the focus here on structural data justice – and more suitable to analyse direct users of data systems than those indirectly affected. Because of that, it might well be more suitable – for those with particular interests in "justice-in-practice" and in justice understood as the expansion of personal freedoms – to take a more capabilities-based approach to conceptualisation, such as those proposed by Heeks & Renken (2017; Figure 3) or Taylor (2017; Figure 2).

D2. Structural Data Justice in Practice

Finally, we can draw from all elements of the SDJ model – including the other aspects of data justice that it encompasses – to propose a "Data-Justice-for-Development Manifesto" as shown in Box 1. This can be used as a guiding set of principles for practice – broad policy; organisational strategy; and specific operational planning – aiming to deliver more just development outcomes from data-intensive development.

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

Cartesian (2014) *Using Mobile Data for Development*, Cartesian, Boston, MA http://www.cartesian.com/wp_content/upload/Using-Mobile-Data-for-Development.pdf

Dencik, L., Hintz, A. & Cable, J. (2016) Towards data justice? The ambiguity of anti-surveillance resistance in political activism, *Big Data & Society*, 3(2), 1-12

Hardy, C. & Phillips, N. (1998) Strategies of engagement, *Organization Science*, 9(2), 217-230

Hearn, J. (2012) *Theorizing Power*, Palgrave Macmillan, Basingstoke, UK

Heeks, R. (2017a) The affordances and impacts of data-intensive development, *ICT4DBlog*, 5 Jun https://ict4dblog.wordpress.com/2017/06/05/the-affordances-and-impacts-of-data-intensive-development/

Heeks, R. (2017b) Data justice: addressing equality, sustainability and transformation through data, presentation at IFIP WG9.4 14th international conference *ICTs for Promoting Social Harmony*, Yogyakarta, Indonesia, 22-24 May

Heeks, R. (2017c) *Information and Communication Technology for Development (ICT4D)*, Routledge, Abingdon, UK

Heeks, R. & Renken, J. (2017) Data justice for development: what would it mean?, *Information Development*, advance online publication

Hilbert, M. (2016) Big data for development, *Development Policy Review*, 34(1), 135-174

Kitchin, R. & Lauriault T.P. (2014) *Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work*. Programmable City Working Paper 2, National University of Ireland Maynooth, Maynooth, Ireland

Kshetri, N. (2014) The emerging role of big data in key development issues, *Big Data & Society,* 1(2), 1-20

Lycett, M. (2013) 'Datafication': making sense of (big) data in a complex world. *European Journal of Information Systems*, 22(4), 381-386

Nambisan, S. (2016) Digital entrepreneurship: toward a digital technology perspective of entrepreneurship, *Entrepreneurship Theory and Practice*, advance online publication

Newman, N. (2015) *Data Justice: Taking On Big Data As An Economic Justice Issue*, Data Justice, New York, NY

Rakesh, V., Heeks, R., Chattapadhyay, S. & Foster, C. (2017) *Big Data and Urban Transportation in India: A Bengaluru Bus Corporation Case Study*, Centre for Development Informatics, University of Manchester, UK

Reynolds, M. (2014) Equity-focused developmental evaluation using critical systems thinking, *Evaluation*, 20(1), 75-95

Sengupta, R., Heeks, R., Chattapadhyay, S. & Foster, C. (2017) *Exploring Big Data for Development: An Electricity Sector Case Study from India*, GDI Development Informatics Working Paper no.66, University of Manchester, UK
http://www.gdi.manchester.ac.uk/research/publications/di/

Spratt, S. & Baker, J. (2015) *Big Data and International Development: Impacts, Scenarios and Policy Options*, Evidence Report no. 163, IDS, University of Sussex, Falmer, UK

Taylor, L. (2017) *What Is Data Justice? The Case for Connecting Digital Rights and Freedoms on the Global Level*, TILT, Tilburg University, Netherlands
http://dx.doi.org/10.2139/ssrn.2918779

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

WWW (2017) *Open Data Barometer*, World Wide Web Foundation, Washington, DC