This is the pre-peer reviewed version of the following article: Heeks, R., Rakesh, V., Sengupta, R., Chattapadhyay, S. & Foster, C. (2020) Datafication, value and power in developing countries: big data in two Indian public service organisations, *Development Policy Review*, advance online publication, which has been published in final form at: https://onlinelibrary.wiley.com/doi/abs/10.1111/dpr.12477

Datafication, Value and Power in Developing Countries: Big Data in Two Indian Public Service Organisations

Richard Heeks[1], Vanya Rakesh[2], Ritam Sengupta[3], Sumandro Chattapadhyay[4] & Christopher Foster[5]

Abstract

Motivation: Datafication – the growing presence, use and impact of data in social processes – is spreading to all sectors in developing countries. But, to date, there have been relatively few analyses of the real-world experiences of datafication in developing country organisations.

Purpose: This paper aims to fill this knowledge gap by analysing evidence of big data in practice in relation to three key issues identified from the literature: implementation, value and power.

Approach and methods: Using fieldwork based on interview, observation and documentary analysis, the paper analyses implementation and impact of big data systems in two public sector organisations in India: one in the electricity sector, one in the transport sector.

Findings: Big data systems have been much slower to implement than anticipated, and the paper exposes the nature and scale of implementation challenge facing such systems. The systems are already delivering value for public service organisations but, as yet, more operational than strategic and incremental not transformative. Big data systems are helping to change profiles of power between and within organisations; for example facilitating a shift in power from public to private sectors, and from labour and middle management to Panopticon-type control by central managers. And big data is intersecting with politics especially around the imaginaries of wider stakeholders, changing their view of the financial and political issues that technology can address.

[1] Global Development Institute, University of Manchester, UK; richard.heeks@manchester.ac.uk
[2] Ex-Centre for Internet and Society, Bengaluru, India; vanyarakesh14@gmail.com
[3] Centre for Studies in Social Sciences, Calcutta, India; rongon86@gmail.com
[4] Centre for Internet and Society, Bengaluru, India; sumandro@cis-india.org
[5] Global Development Institute, University of Manchester, UK; christopher.foster-2@manchester.ac.uk

Policy implications: Policy-makers and practitioners can better understand and plan for the role of big data in development using three frameworks presented in the paper: the information value chain, the decision pyramid, and the big data—power model. These expose key issues of implementation, value and power that must be incorporated into big data policy and projects, and that lead to longer-than-expected timescales. They also shape an agenda for improved knowledge-sharing and capacity-building around big data; actions that will address both practical and political issues arising in the datafication of development.

Keywords: big data, datafication, India, public sector, value, power, politics

Acknowledgements
Funding for this research was provided from a joint fund of the Universities of Manchester and Sheffield.

Introduction

Datafication may be defined as the growing presence, use and impact of data in social processes. Particularly impelled by the diffusion of digital technologies, it is spreading to all sectors within developing countries (Taylor & Broeders 2015). There are examples of researchers using new datasets to demonstrate the potential contribution of data to development (Albanna & Heeks 2018). And there are speculations, hopes and concerns about the datafication of development (Spratt & Baker 2015, Hilbert 2016). But, to date, there have been relatively few analyses of the real-world experiences of datafication in developing country organisations.

This is particularly true of big data. This is a label typically attached to new datasets that represent the "3Vs": much greater volume of data than previously available, but also greater velocity (i.e. reduced time lag between an event and availability of data about that event) and greater variety (i.e. integrating more diverse forms and types of data) (Dumbill 2012). The implementation of big data systems in developing countries is still quite limited and quite recent. As a result, there have not been many studies – literature has been based mainly on prototype applications and/or conjectures – and the purpose of this paper is to address this knowledge gap. It selects two case studies of big data application in the Indian public sector; one in the electricity sector, one in the transport sector. It then asks how these were implemented and what impact has been associated with these applications; both the organisational value of big data and also the wider intersection between big data and power. The paper's main contribution is to provide a systematic insight into the challenges, impacts and implications associated with datafication of development based on substantive real-world case evidence.

The paper is divided into five sections. Next, we will review the literature to date on datafication and development, looking specifically at the role of big data. We will identify frameworks that can used for analysis of datafication and development. We then describe the two Indian case studies, and our fieldwork that gathered evidence on those cases. The findings of the paper are divided into three parts. The first examines implementation of big data systems. The second examines the organisational value so far being derived from big data. The third analyses the shifts in power associated with big data, and also the way in which big data intersects with wider politics. Finally, we draw some overall conclusions.

Big Data in Developing Countries

We base our analysis of evidence to date on big data and developing countries on nine existing literature reviews on this topic (Kshetri 2014, World Bank 2014, Lokanathan & Gunaratne 2015, Spratt & Baker 2015, Taylor & Broeders 2015, Ali et al 2016, Hilbert 2016, Sengupta et al 2017, Albanna & Heeks 2018). We analysed this literature in terms of a foundational model of data systems: the information value chain (Heeks 2018). This process model charts the chain of steps linking a data system to development results: capture of data, its processing and visualisation as information, the use of that information to inform decisions, which lead to actions, which create results. Shaping most of those steps is a framework of knowledge which can itself be developed through addition of new information.

Viewing the literature on big data and development through this model, we were able to identify three main types of issues within that literature; related to implemention, value, and outcomes (see Figure 1).

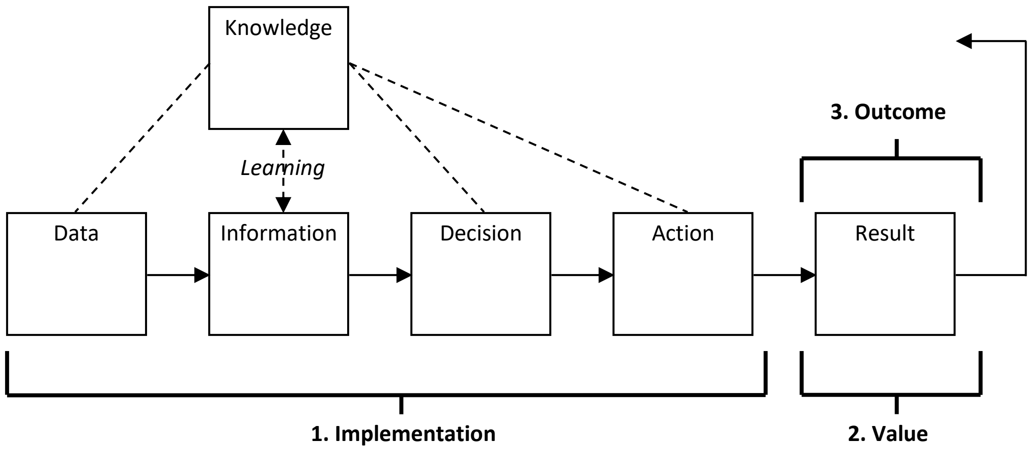

Figure 1: Issues in Big Data and Development Literature Mapped to the Information Value Chain Model

The literature covering big data *implementation* focuses on challenges that either have to be overcome in order for big data systems in developing countries to operate, or which are preventing fully-effective implementation of those systems. Two domains of challenge dominate in the literature. There are those relating to technical issues particularly in the "upstream" parts of the information value chain; i.e. linked to handling of data and information. These relate to digital technology such as lack of devices to capture data, lack of telecommunications infrastructure to reliably transmit data, lack of computing power to analyse data, etc. (Kshetri 2014, Ali et al 2016). They also relate to data itself: the limited extent of digital data in developing countries, its relative inaccessibility, its poor quality, its fragmentation across different sources, etc. (World Bank 2014, Lokanathan & Gunaratne 2015, Ali et al 2016, Hilbert 2016, Albanna & Heeks 2018). And they relate to the need to develop new techniques in order to analyse and visualise development-related datasets (World Bank 2014, Lokanathan & Gunaratne 2015, Sengupta et al 2017). The second domain of challenges relates to human capabilities. Again, discussion tends to focus on the upstream parts of the value chain. So there are concerns about developing countries lacking enough skills to clean, integrate, analyse and visualise data (Kshetri 2014, Lokanathan & Gunaratne 2015, Hilbert 2016, Albanna & Heeks 2018) more than concerns about the capacity to make decisions and take actions on the basis of that data.

Value deals with the direct organisational and developmental improvements that big data can make. The literature reflects that there are relatively few examples as yet of big data application leading to large-scale, sustained development results (Taylor

& Broeders 2015, Albanna & Heeks 2018). Thus, most of the examples cited in the literature to date are what might be called "proof-of-concept". These are projects – often led by computer/data science researchers – that show the potential for a contribution but not an actual development impact. For instance, demonstrating the feasibility of using social media to track food prices (UNGP 2014 cited in World Bank 2014) or demonstrating the potential to measure poverty from mobile phone call detail records (Blumenstock et al 2015 cited in Albanna & Heeks 2018). There is also more general reflection on the potential of big data to contribute to global development, relying heavily on the proof-of-concept literature or on experiences in the global North (e.g. Lokanathan & Gunaratne 2015, Spratt & Baker 2015, Hilbert 2016).

Because of this early-stage, overview nature of evidence to date, conceptualisation of the value of big data for development has been very limited. In the main all that has been offered is high-level categorisation of big data applications. These are categorised by development sector (Lokanathan & Gunaratne 2015, Ali et al 2016, Albanna & Heeks 2018) or by type of data (World Bank 2014, Ali et al 2016, Hilbert 2016, Albanna & Heeks 2018). But this offers no detailed insight into value. For that, one needs a micro-level perspective, drawing from the information value chain's indication that the value of data comes from its conversion into information and use in decisions.

A generic approach links the value of big data to the features of the data, in enabling decisions which are quicker and/or lower-cost and/or of better quality than was previously possible (Kshetri 2014, Albanna & Heeks 2018). But that offers no differential sense of value. For the analysis that follows, we therefore adopted one of the most widely-used models of organisational decision value: the pyramid notion of three levels (e.g. Turban et al 2018). Operational-level decisions involve day-to-day supervisory actions dealing directly with front-line employees. Tactical-level decisions relate largely to middle-management, medium-term processes. Strategic-level decisions are those which relate to the overall goals of the organisation. As well as relating to types of decision, this model also represents levels of value: it is strategic decisions which can deliver the greatest value.

Outcome issues are the broader impacts associated with growing use of big data in developing countries[6]. These are increasingly being analysed through the lens of "data justice" (Heeks & Renken 2018) from which two main perspectives emerge from the literature. The first relates to data rights. For example, every one of the literature reviews mentions fears about loss of privacy either through illegal access to big datasets or through growth of surveillance of populations by the state or by private corporations. The second relates to data-based inequalities. Most of the reviews implicitly or explicitly relate this to a "big data divide" (Hilbert 2016). This generally means that unequal access to digital technology or to big datasets or to

[6] Value can be associated with the idea of "imminent development": the intended, incremental changes arising from application of big data; while outcome can be associated with "immanent development": the wider or longer-term, sometimes unintended changes that are emerging as big data is applied in development (Hickey & Mohan 2005, Murphy & Carmody 2015).

data-related capabilities leads to unequal access to the development benefits of big data. And it also means that lack of representation of some populations within big datasets means they are denied the benefits of decisions and actions that use those datasets.

Some of the reviews (e.g. Kshetri 2014, Spratt & Baker 2015, Hilbert 2016) recognise that these issues may empower some actors such as large corporations or the state at the expense of others. But only Taylor & Broeders (2015) explicitly link these inequalities to concepts of power, noting the lack to date of – and need for – power-based analyses of big data and development. They identify a shifting terrain of power; particularly a potential "privatisation" of development if power shifts to those private sector actors who control big data resources, capabilities, discourse, etc. Building from this work and from Heeks (2017), we can develop a simple model which can be used to analyse datafication and power. This identifies a number of sources of power that can impact the central practice of a data system; understood as the information value chain. These sources include:

- control over resources: specifically who owns and controls the data within the big data system;
- epistemic control: in terms of who has the knowledge developed by the data system, and the impact of the wider discourse around the system;
- institutional control: understood in relation to both formal institutional forces like rules and regulations, and informal institutional forces like norms and values;
- structural position: the implications of new structures and positionality of relations arising in association with the big data system.

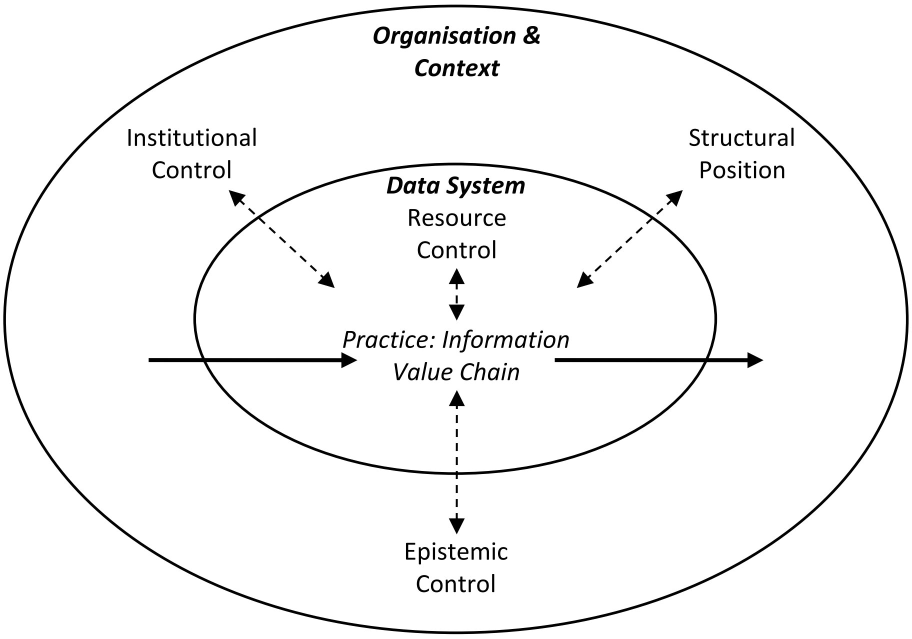

Figure 2: Big Data Systems and Power

Overall, the emergent literature highlights a set of key issues in application of big data to development. There has been discussion of the different technical and human challenges that may arise during implementation. However, with the lack of scaled cases, there is little exploration of how implementation links to changes in value within organisations. Likewise, while there has been some hypothetical discussion regarding power and data, there is little instantiation of this in developing country contexts. We therefore now move to apply the models identified in this review to the two big data case studies; described next.

Case Background and Methods

In order to investigate these issues further, we undertook two case studies of big data-based datafication in India[7]. The first involves "Stelcorp": an anonymised state-owned electricity generation and distribution corporation in "Janakari" state. The second is the Bengaluru Metropolitan Transport Corporation (BMTC) which runs buses in Bengaluru. Following an initial scanning exercise of possible cases in India, these two were selected because their experiences with big data – while still relatively recent – had progressed far enough for us to draw conclusions about implementation, value and outcome.

The findings below were drawn from four sources which were combined and then analysed through the lens of the models presented above. First, seventeen interviews with a mix of senior, mid- and operational-level staff in both organisations, including some technical staff. The interviews also included discussion with four external technical contractors and advisers. Second, direct observation of the big data systems in operation, at the stages of data capture and visualisation and use of that data. Third, a set of primary documentation: tender documents, technical reports, strategic plans, annual reports, audit/regulatory reports, policy documents, trade union correspondence and newsletters, etc.[8]. Fourth, secondary sources of direct relevance such as published case studies from private partners, evaluation reports, case analyses as well as academic literature on both organisations.

Case One: Stelcorp

Stelcorp, the state electricity generator, serves more than 15 million customers, of whom around 80% are in rural areas, employs more than 15,000 staff, and has revenues in excess of US$1bn per year. It works with a number of private sector ICT partners including a main firm, "Digicorp". These ICT collaborations have been particularly aimed at addressing the problem that, in India, an average one-third of power generated is lost (Min & Golden 2014). Losses fall into two types. Technical losses result from resistive/heat dissipation of energy as it is transmitted and from breakage or malfunctioning of grid components. But the majority of losses are commercial, including "inefficiencies in billing, meter tampering, illegal connections, and use by flat rate consumers that exceeds their nominal allotments" (*ibid.*:619).

[7] Findings in this paper integrate material from two working papers that describe each of the case studies in greater detail: Sengupta et al (2017) and Rakesh et al (2018).
[8] The requirement for anonymity in the case of Stelcorp has meant it is not possible to cite some of these documents since to do so would enable case identification.

In 2008, and building on earlier programmes, the central government initiated the Restructured Accelerated Power Development and Reforms Programme (R-APDRP) (Narasimhan et al 2011). This provided funding for digital information systems to address both technical and commercial aspects of loss. In Stelcorp's case, total expenditure of around US$50m has been used to invest in a series of new technologies (see Figure 3):

- Automated, online digital meters were installed across the higher levels of the power network, especially in urban areas. These monitored all urban feeders that step-down voltage between sub-stations and transformers (typically serving around 25,000 customers); all urban transformers that step-down voltage to household levels of 230V (covering 50-100 customers); and all bulk power consumers (such as factories, food processing plants, cold storage facilities, etc.). These capture and transmit a variety of network "health parameters": voltage, current, load, energy and event alerts such as tampering or power outages.
- Stand-alone digital meters have been installed on all rural feeders and transformers. These perform the same function as the online meters but they must be read manually by a human meter-reader. They, in theory on a monthly basis, offload data onto a handheld device – the Common Meter Reading Instrument (CMRI) – and then upload that data at their divisional office.
- Basic digital meters that just record level of energy usage have replaced almost all of the pre-existing analogue meters for ordinary customers. That data must be entered into the CMRI by a human meter-reader and then uploaded at a local centre largely for billing purposes.

All of this data is fed into an organisation-wide data system: the Meter Data Acquisition System (MDAS). ICT partner Digicorp developed this system, and is also responsible for designing and operating the software architecture underpinning MDAS, for data management, and for training of Stelcorp staff. MDAS acts as the organisational data repository, now storing hundreds of terabytes of data on energy use, faults and losses, plus also data on revenues. It can process and communicate that data across the organisation since MDAS is available in all of Stelcorp's administrative and technical offices.

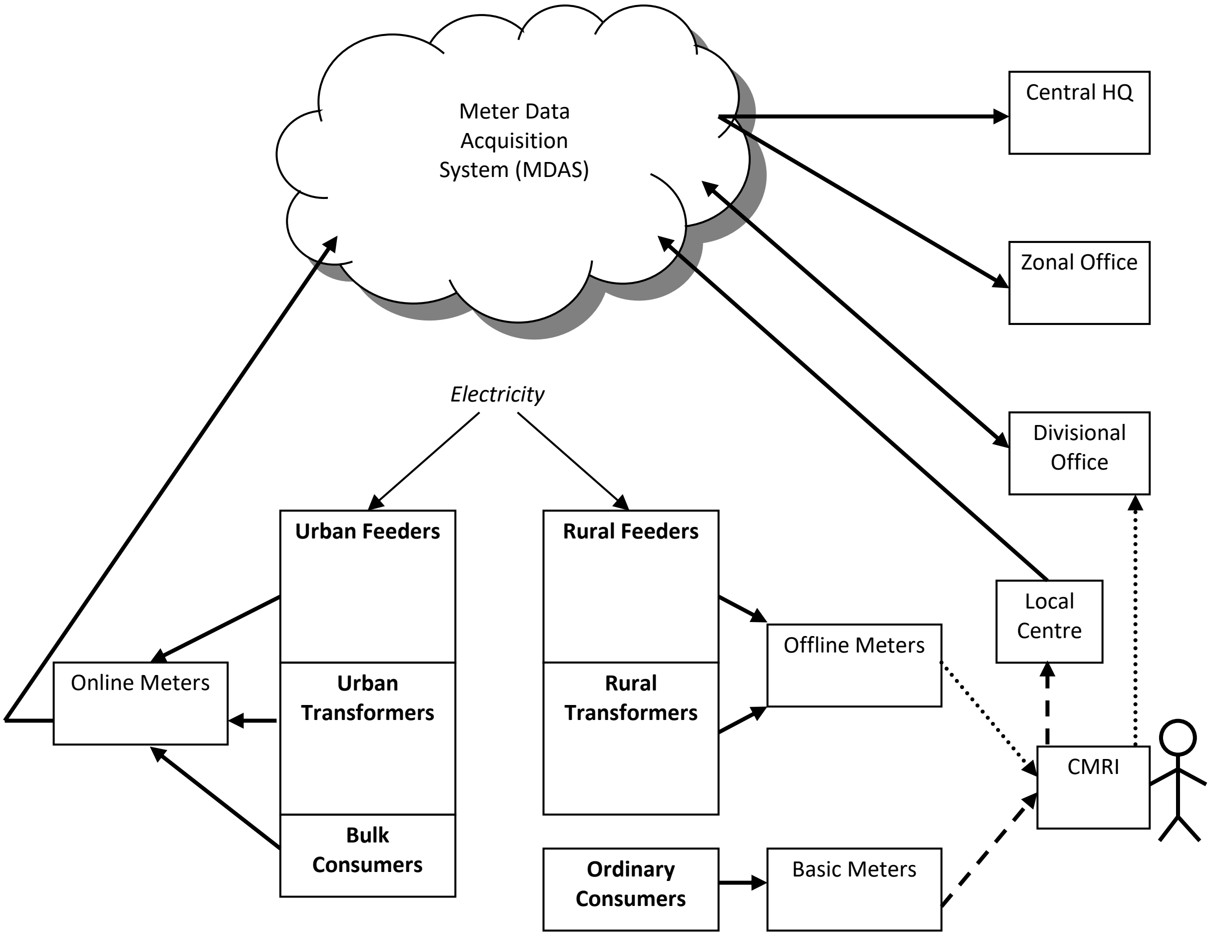

Figure 3: Stelcorp's Data System

Case Two: Bengaluru Metropolitan Transport Corporation

The Bengaluru Metropolitan Transport Corporation has more than 6,000 buses and nearly 35,000 employees. Every day it carries more than five million passenger journeys over more than 2,000 routes and 8,000 bus stops in Bengaluru, collecting over US$500,000 in revenue, and its buses travel more than 1.2 million kilometres (BMTC 2018a, BMTC 2018b).

Such an operation requires a large number of operational, tactical and strategic decisions, and generates huge amounts of data which could be used to guide that decision-making. Historically, that data has not been captured or has been captured very inefficiently using human-based or paper-based systems. With growing size and complexity creating increasing decision-related problems, in May 2016, BMTC launched its Intelligent Transport System (ITS); a project initially conceived in 2012, beginning implementation in 2013, and with an initial cost of around US$1.2m (Akshatha 2015). The main implementer was private sector IT company, Trimax, though the Corporation was also advised by two private sector consulting firms.

ITS involves four main components (see Figure 4):

- Vehicle Tracking System (VTS): installed in all 6,400 buses, this has a tracking unit that uses the city's mobile telecommunications network to transmit real-time data on the location of the bus. It also includes a "voice kit" that allows two-way communication between the bus crew (driver plus conductor) and either the BMTC central Control Room, or managers at the bus depot (the major bus stations, of which there are just over 40 in the city).
- Electronic Ticketing System: there are more than 10,000 of these handheld-devices which, resembling a credit card reader with integral printer, are used by bus conductors to issue tickets. They record ticket price and time alongside bus stop of issue, bus number, etc. This data is then uploaded to the central system every five minutes via the mobile network.
- Central Information Systems: the main servers record up to a hundred gigabytes of data every day, and allow that to be visualised in the Control Room which contains a wall of large screens. Data is also routed into a Depot Application, which can generate specific information such as crew duty rota, the log sheet for the driver, kilometres travelled and details about fuel usage, ticketing, etc.
- Passenger-Facing Systems: data from the Central Information Systems feeds a set of passenger interfaces that provide general timetables, maps, journey planners but also real-time updates. These include overhead display screens and touch-screen kiosks at major bus stations, and a mobile app.

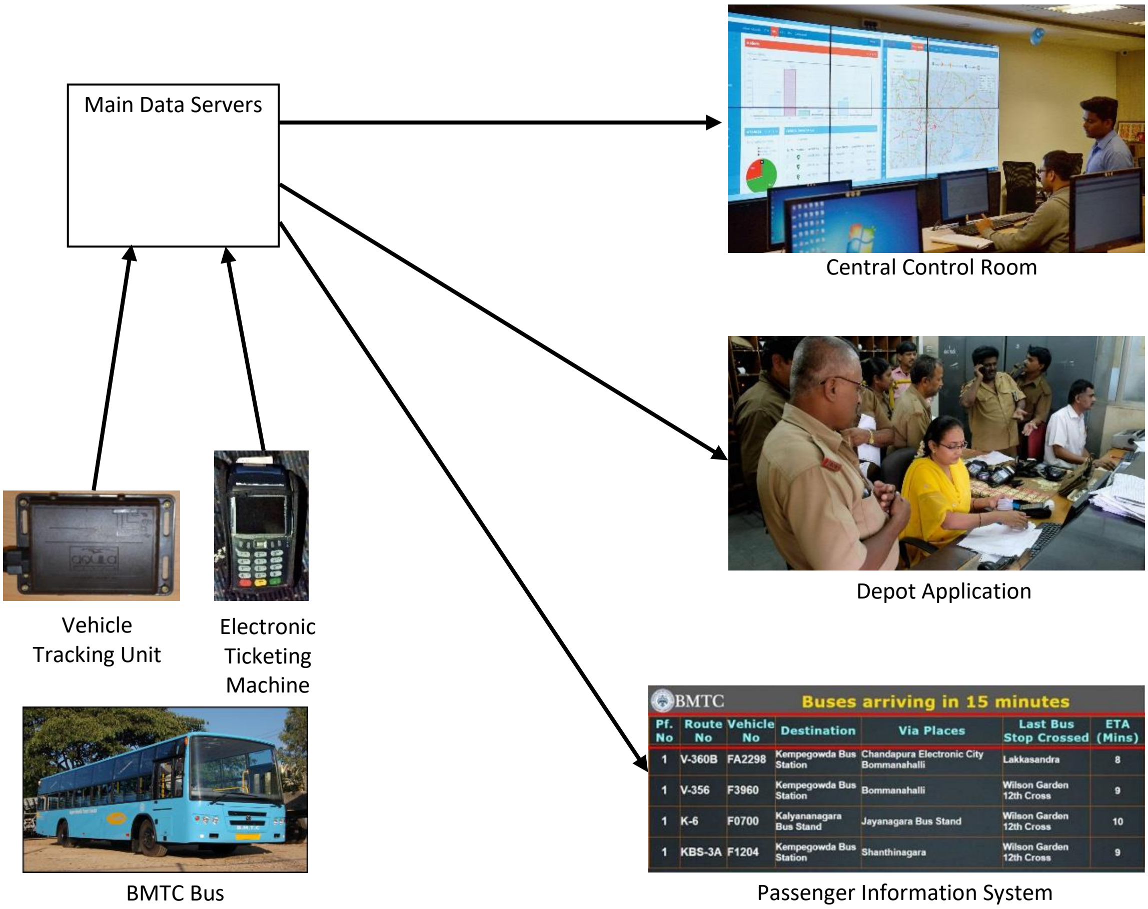

Figure 4: BMTC's Intelligent Transport System
(Photo sources: BMTC 2013, Bhattacharya 2016a, BM 2016, Ramakrishnan 2016, BMTC 2017, Jha 2017)

Findings

Implementation

Both applications have taken some years to get the data component of the information value chain functioning effectively; something which is a necessary foundation for the rest of the chain to operate. A number of the issues have been technical. For example, in terms of equipment, Stelcorp engineers complained about the functioning of the meters, with some claiming that up to half of meters were unreliable in terms of their recording or connectivity; thus undermining the availability and quality of data. In BMTC, there had been initial difficulties with breakdowns in mobile network connectivity, and with the ticketing machines running out of charge. Both of these had required system design revisions but managers still reported connectivity problems at times. The passenger-facing systems have also had ongoing problems: while they do undoubtedly work at times, fieldwork visits to depots found neither public screens nor kiosks to be operational. Similarly the mobile app has had many complaints about bugs, slow operation, and limited abilities to plan routes or predict bus arrival times (CM 2017 and customer reviews on Google Play store).

A second technical issue has been the data foundation: for big data systems to work, they need to digitally and uniquely identify very large numbers of organisational assets. For BMTC, it took many months to map and identify its 8,000 bus stops, many of which had the same name (Jha 2017). Stelcorp's challenge was far greater – to geo-locate all of its more than four million assets (feeders, transformers, pylons, etc.) and issue a unique ID to all of its more than 15 million customers. This took five years and, though declared completed in 2015, interviewees stated that some errors in both geo-location and customer ID coding regularly emerge.

There have also been technical challenges around interoperability: big data necessarily means bringing together many disparate elements of an organisation, and this creates problems which were seen particularly for Stelcorp. Engineers reported that meters have been sourced from several different producers, which has made it hard to standardise data into a common readable format: some still has to be re-assembled by human intervention. There has also been a problem linking the new billing information system and MDAS. The billing system was designed on the basis of a common meter reading and billing cycle for all consumers linked to a specific transformer; i.e. everyone billed on the same day each quarter. But both practice and the data system in Stelcorp are based around different meter reading and billing dates for all consumers with no common cycle. This has left some customers unbilled, and has a knock-on for accuracy of calculating transformer-level losses.

Other challenges have been human. Both organisations have had to undertake major training initiatives to build the necessary skills for operation of the new technologies and data systems. But reported challenges were less about skills than about acceptance of technology. At BMTC, drivers and conductors were concerned about the surveillance potential of the new devices.   So, there were a few examples of the vehicle tracking units being burned, filled with water, or shielded with metal casings in the early days of implementation (Jha 2017). At Stelcorp, managers indicated that a small minority of human meter-readers were not going out to capture data direct from the digital meters but were, instead, just logging average values based on consumption levels in earlier months. In both organisations, interviewees and other sources (e.g. Bhattacharya 2016b) indicated there had been more significant resistance to introduction of the new systems from middle managers.

Value

The implementation challenges identified have, over time, been sufficiently overcome that the information value chain is operating at least to some extent. The data collected is being processed into information, which is feeding decisions and actions in both organisations, leading to organisational results and value. We can now analyse this in relation to the three types of decision identified earlier: operational, tactical, strategic.

At the *operational level* in the bus corporation, all those interviewed and all reports indicated that the vehicle tracking units and electronic ticketing machines were now working well enough to allow new decisions and actions. For example, analysis of driving patterns from the VTS data provided for exception reports to be triggered in the central Control Room: skipping of bus stops, deviation from planned route, speeding, or being stationary for a significant period. Control Room staff reported and could be seen then intervening with individual bus crews on the basis of those reports, investigating issues via the two-way voice kit.

In the bus depots, managers reported two main impacts (see also Jha 2017). First, introduction of the ETM had made crew roster scheduling, end-of-day audit and cash reconciliation easier. This had led to a reduction in revenue discrepancies, and a reduction in the number of cases of "ticketing pilferage" though it was unclear if this covered theft by bus crews. Second, traffic police would previously issue fines – which the depot manager had to handle – for claimed traffic violations by bus drivers. And other road users would issue compensation claims for accidents. In a number of cases these were known to be fake but there was no mechanism to check whether a bus was in fact speeding, or was stopped as a result of an accident. Though imperfect and not providing all necessary data (e.g. it cannot unequivocally tell if a bus has jumped a red light), data from the Vehicle Tracking System allowed some ability to cross check. As a result – perhaps from the perception as much as the reality of vehicle tracking – the number of fake fines and compensation claims issued was said to have significantly reduced.

Tactical-level decisions relate particularly to management of routing, scheduling and fare setting. At the time of the case study, these uses of big data appeared to be more potential than reality. For example, BMTC plans to use ticketing data to identify routes with low passenger numbers during off-peak hours and to offer reduced off-peak fares on those routes. Likewise, there were aspirations among BMTC interviewees to use the data to rationalise routes, schedules and bus operations in order to better respond to daily and seasonal patterns of demand for public transport, and to enhance efficiency in the operation and management of crews. But these plans had not actually been implemented.

One may also judge ITS at the *strategic level*. As yet, there is little sign of contribution to broader urban transportation goals. We can judge on the basis of the four BMTC goals to which ITS was intended to contribute (Mishra 2016, BMTC 2018b, Kidiyoor 2018):

- Service reliability: 15% of services were cancelled in 2017-18 compared to an average 6.5% in the five years prior to ITS introduction.
- Citizen centricity: the passenger-facing systems including the mobile app have been steps forward but they are not fully operational or effective as yet in providing information and journey planning support for travellers. Passenger-journey numbers were down by about 500,000 per day in 2017-18 compared to

2015-16, so aspirations of helping to move more passengers onto the buses were not yet realised.

- Revenue management: there is little internal data on this but overall passenger revenue fell by 8% from 2015-16 to 2017-18 compared to an average 10% growth over the previous five years.
- Organisational capabilities: fleet utilisation fell below 90% for the first time in 2016-17 and was lower still in 2017-18 while costs per km were at their highest-ever level.

There are of course many internal and external factors other than ITS that contribute towards organisation-wide performance; not least the steady extension of the Bengaluru Metro system which carries passengers who might otherwise use the buses. However, this does show a lack of any overt short-term impact of ITS on strategic performance indicators. It also suggests a lack of contribution as yet to even wider measures of urban transportation relating to journey times, congestion, pollution, etc.

Given its longer history, it is not surprising that big data in Stelcorp is providing more value and contributing across the range of decision-making levels. *Operational* activity focuses on billing and on dealing with faults in the electricity distribution system. In general, billing was reported to be more accurate than previously but it is not faster since it occurs via the same, i.e. quarterly, cycle times as before. Faults – short-term power outages caused by equipment malfunction – used to be detected via laborious trial-and-error testing in the field by linesmen[9] based on a mix of guesswork, intuition and experience. Faults are still not reported automatically by MDAS but mainly continue to be reported from outside by customers or linesmen. However, managers can now drill-down within MDAS to identify likely location of the problem, and related data on the functioning of relevant online (i.e. excluding rural areas) assets like transformers. Linesmen are thus more efficient in finding the fault source. A "docket" system also operates for each fault that automatically identifies to managers faults not rectified within a set time frame. Thus, alongside better and faster location of problem sources and transmission to field staff and resolution, there is also better monitoring of the process.

Tactical decisions relate to loss detection and internal performance management. Unlike faults, losses are identified on a longer-term basis by retrospective calculation and may have multiple causes which could be technical or human. The presence of MDAS has made it easier and quicker to identify the source of losses. As with faults, the information value chain still operates in "pull" rather than "push" mode: managers must find relevant data within the system rather than being alerted via automatic exception reporting, and their pull is reactive – driven by external events – more than proactive identification of problems.

Regarding internal performance management, there has been a structural innovation within Stelcorp, with the formation in 2015 of the central Finance and Energy Management Unit (FEMU). FEMU has responsibility for oversight of all

[9] They are all men.

organisational data including performance of the distribution network and its components including losses, and adherence to consumer billing cycles including levels of collection and disconnections. These can be broken down into the various zones and then districts of the Corporation.

These performance indicators then lead to internal decisions and actions. FEMU holds general monthly meetings with zonal- and divisional-level managers to review performance and plan interventions e.g. further actions on loss prevention. It also holds weekly video-conference discussions with individual offices with the same function: review of past data and planning of future actions. This is something that officers complain about. Historically, they were accountable to their immediate superior and would discuss, explain, negotiate and agree performance issues with that person. That has now been replaced with direct accountability to FEMU, with whom – given FEMU's whole-organisation span of control – relations are more distant and formal; an outcome with which many managers are uncomfortable.

MDAS also feeds into more *strategic* planning such as electricity network developments. Plans for major extensions or upgrades are based on a mix of three levels of data: top-down political / policy directives; bottom-up complaints or demands from consumers; and "mid-across" historical technical data from within Stelcorp. While all three are used for selection of locations for action, the latter is used for specific planning of initiatives. It has been historically weak, based on basic, aggregate statistics and fairly rudimentary calculations. Now, though, there is more objective data to stand alongside the relative subjectivities of politics and consumer complaint. Modelling of upgrades to the existing network can now be based on more quantified, objective data, though it still appears to be largely manual or IT-assisted rather than itself being automated. Modelling of extensions into entirely new areas will only have incrementally improved because by definition there is no MDAS data for as-yet-unserved areas; though related data will be used. One sign that this is not so robust as the data and modelling used for upgrades was an audit review into Stelcorp's rural electrification programme which found one-quarter of newly-installed transformers had failed due to load under-estimation.

Likewise, and as with the BMTC case, if one steps back to the broadest scope of value, the picture is disappointing. The whole intention of the new data systems was to reduce losses and to make associated improvements such as increasing revenue collection efficiency. In the years since 2009, this has not happened. Taking three key parameters:

- Aggregate technical and commercial (AT&C) losses – the overall measure of difference between units of power supplied and revenue collected – dropped significantly in 2010/11 but then rose back up to earlier levels and has remained roughly static at one-third up to 2017-18. This, as noted earlier, has been a typical average across India.
- Billing efficiency – the proportion of supplied electricity that is billed to customers – was static until 2011/12 but has dropped since by a few percentage points to around two-thirds.

- Collection efficiency – the proportion of billings collected – has been volatile between 93% and 100% throughout the period since big data was implemented without so far demonstrating stability of improved performance.

As discussed later, there are reasons for this but, more than five years after the major big data-related reforms began, their impact is not yet seen on these most-aggregate measures of Stelcorp's performance.

Outcome: Power and Politics

Big Data-Associated Changes in Locus of Power

In both cases, though particularly Stelcorp, the advent of big data has been associated with shifts in the locus of power of which we will consider four here.

First, there has been an upward shift in power from labour to management. Central managers now have greater resource and epistemic control through the Panopticon effects of the big data systems. For the bus crews, the central Control Room can directly see in real-time the behaviour of drivers, and can also call up details of conductor ticketing. Bus crews no longer control that data or the knowledge of their behaviour that entails. It has also allowed practices to change, with managers directly intervening to administer bus crew actions. Likewise, for meter-readers, central managers can now directly monitor their behaviour and performance, and make decisions accordingly. The technologies associated with big data systems can also bring automation of human labour; thus reducing the importance of labour as a resource in the organisation. In the years since roll-out of the system, 40% of meter-readers lost their jobs due to introduction of online meters. Continuing roll-out will cut their numbers even more. For bus crews, the partial automation of ticketing has eroded the role of and necessity for two-crew operations, and pressure for substitution of the conductor's role with digital technology will continue to grow.

The second shift is that from middle to central management. Previously, data was relatively "sticky"; tending to be restricted to localised enclaves within the organisation. Staff would thus be accountable to their immediate middle managers. It was those managers who accessed data on staff, negotiated its implications, and held power over their juniors. Big data systems enable central management in BMTC and Stelcorp to directly access performance data from all levels of operations. New practices are underway by which they undertake and communicate performance management, obviating the need for intermediating management layers. Institutional power of authority is shifting as staff and managers now significantly see themselves as accountable to BMTC's central Control Room and to FEMU in Stelcorp, and a looser institutional power is arising as big data is allowing a stronger audit and performance management culture. In terms of positionality, the central units now have a direct structural relationship to organisational staff and managers. As a result, both organisations have begun stripping out middle management layers; reducing numbers and, in the case of BMTC, closing five divisional offices that used to intermediate between the depots and the central management team.

The third shift is that from public to private sector. Previously, the corporations were responsible for their own data and data systems. This is no longer the case. While *de jure* ownership lies in the hands of the corporation, *de facto* ownership has shifted, particularly in the case of Stelcorp. The data system providers – Digicorp for Stelcorp, Trimax for BMTC – have control over a key resource: knowledge of the design, construction, operation and maintenance of the data systems. While they could in theory be replaced, it would be very difficult to do this in practice due to the criticality of the data systems and dependency on the private providers' knowledge. Hence, for example, complaints from Stelcorp staff. Engineers were unhappy about the speed with which Digicorp fixed data system problems, but felt they could to little about this. IT staff complained about their lack of control over Stelcorp data and software, their dependency on Digicorp, and their diminution to what they saw as just a clerical role. Managers also raised concerns that Digicorp could make and possibly is making commercial use of analytical value from the Stelcorp data; for example selling data or analysed information to electrical appliance and equipment manufacturers.

Finally, one can note a shift in the locus of power that has not occurred much: a shift from the public corporations to citizens and citizen groups. In the transport case, there is some greater operational transparency through the – albeit flawed – passenger information systems. But neither corporation has embraced open data[10]. Talking with managers, there was no open data impetus within Stelcorp and the only data available in the public domain is rather outdated aggregates like division-wise AT&C losses. For some time, BMTC has been talking about plans to open up its real-time data (e.g. Akshatha 2016), but at the time of writing there was no visible sign of this.

Big Data and Politics

There is a symbolic, interpretive aspect to the epistemics of big data (Boyd & Crawford 2012). Big data changes the "imaginary" of organisations: not just what is visible about the organisation, and to whom, but also their perceived image of the organisation (Boellstorff 2013, Taylor & Broeders 2015). For example, in BMTC, 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. Big data is thus changing the landscape of what is seen to be possible within the organisation. For example, there has been a revival of plans for driver-only buses – a change trialled in the 1990s but then abandoned – but now once again seen as possible by managers; something they specifically link to the advent of ITS (see also Philip 2017).

This is seen even more strongly in Stelcorp's case. As noted above, overall losses have not reduced and the percentage of customers who receive bills has actually gone down somewhat since the advent of big data. This has happened because

[10] The potential 4th V of big data: increased visibility (Heeks 2018).

technology has intersected with the politics of electrification in Janakari State. Respondents, internal reports and reports of the Electricity Regulator all aver that the big data innovations described above are reducing technical losses from transmission resistance/heat and from faults. There was also some evidence that online metering was reducing losses from bulk consumers in urban areas. But there has been a compensatory increase in theft and defaulting among ordinary consumers; especially those in rural areas.

This is associated with what interviewees described as Stelcorp's "unprecedented" expansion of rural connections in recent years, reflected in the doubling of the consumer base. It is the newly-electrified rural areas of the state which have particularly been returning highly-inefficient billing patterns and high loss levels: above 50% compared to losses well below 25% in urban areas. Those high loss levels arise due to the politics of electrification. Rural electrification has been a significant electoral issue in the state and a political tool, with parties seeking support by ensuring maximum access to electricity for rural constituents with minimum imposition of payment[11].

This leads field-level staff into direct conflict with local politics as they seek to collect unpaid bills or disconnect defaulting or illegally-connected consumers. The result has been repeated complaints from staff trade union representatives about growth in theft of electricity in rural areas, about an increase in threatened and actual violence against field staff, and about a lack of effective response from either Stelcorp or the local police. In some cases, Stelcorp officials rationalised all this as a "social cost" that utilities had to pay, and which still contributed towards national development goals.

One reading would see this as a silencing of big data in the face of big politics: the capability for accurate metering and billing of almost all consumers simply being overridden by electoral imperatives. But there is a little more to it than this. From the comments of regulators and Central Government, it becomes clear that the reductions in technical and urban losses thanks to big data are necessary in order to allow the State Government to "get away" with its political approach to rural electrification. The technical and urban loss reductions offset the growth in rural losses and so Stelcorp avoids regulatory intervention when reporting overall performance to national regulators.

Second, big data has changed the epistemics of electricity; has changed in particular the political imaginary of electricity in the state. Interviewees explained that, previously, the mental image politicians and other senior officials had was one of blackouts and brownouts, requiring slow, limited, and incremental spread of the electricity network. Big data seems to have changed that – partly through the reality of improvements it is delivering such as faster fault rectification but equally through the promise that it holds – to a paradigm of continuous, high-quality, universal

[11] See Min & Golden (2014) for analysis of the wider pattern in India of rising commercial losses in the run-up to state assembly elections as politicians seek to attract votes by interfering with the electricity supply process.

power. Through the transition of imaginaries partly facilitated by big data, policy and strategy have therefore changed; feeding into and being fed by the ambitions of politicians to deliver electricity everywhere and everywhen. New plans and promises are now based on the idea that all districts – and all voters – can have 24 x 7 power. In a way, then, the data-enabled growth in connectivity has helped fuel a politically-enabled growth in free appropriation of electricity.

Discussion and Conclusions

This final section will first review the findings; once again using the three-issue structure derived from the literature: implementation, value and outcome.

Following multi-year ***implementations***, both of the big data systems are operational in their corporations, based around fast-growing datasets already involving tens up to hundreds of terabytes of data. As discussed in the literature, there have been a series of challenges facing these systems. Most of these can be mapped against the upstream parts of the information value chain: sorting out foundational datasets, and the operation and use of equipment from which ongoing data derives.

Much of this would be anticipated, given that these are the first and large-scale applications of big data in these large public sector organisations. They suggest that time-scales for big data implementation will be several years, even to get to a basic operational state. Second, that even when operational, there may be an ongoing requirement for human intervention in parts of the system that are intended to be fully automated. Most of these issues are likely to be transitional rather than permanent: like a number of the findings, they reflect that we are still relatively early in the life-cycle of these systems. However, some of the problems have been persistent and they not ephemeral: instead, they appear inherent in moving to big data-based operations.

The challenges appear more technical than human-related. This is somewhat surprising. The literature expresses concerns about lack of big data capabilities in developing countries but the cases did not really reflect this. Most likely this was because the organisations had had several years to build capabilities and because, as discussed below, they were not yet fully exploiting all the potentials of big data and were relying on capabilities within their private sector partners. Resistance to big data systems was not a significant theme in the literature and, while observed, it was somewhat muted. Both Stelcorp field staff (linesmen, meter readers) and BMTC bus crews are heavily unionised, and the big data-related technologies represent a potential or actual threat to jobs. Yet resistance to these technologies has been individualised and fragmented. Union grievances and strikes have related to traditional issues such as pay, assaults on staff, suspension of staff; not to digital technologies. Instead, greater resistance has come from middle managers in both organisations, as they seem to have more-clearly recognised the threat that big data systems pose to their jobs.

That resistance, alongside the much broader challenges of organisational change in large, public sector organisations explains two other patterns seen in relation to the extraction of ***value*** from big data. First, that it was taking some time for value to emerge. In terms of the information value chain, it is upstream processes of datafication that have appeared more readily than downstream conversion of that data into decisions, actions and results. Second, that – while those decisions, actions and results are now definitely in evidence – a relatively consistent pattern is that the direct impact of datafication is being felt most in relation to operational decisions; least in relation to strategic decisions; with tactical decisions somewhere in between.

From this, it can be argued that, to date – while these organisations have moved far beyond "proof of concept" – they have only extracted relatively incremental value from big data. Changes have occurred most at an operational level to the work of front-line staff. There is no sign yet of the value of big data in the most strategic, organisation-wide performance indicators including the major issues such as electricity losses in Janakari state and traffic congestion in Bengaluru. The pattern of incrementalism can be seen for each of the 3Vs of big data. The volume of data reflects capture across the whole breadth of each corporation's operations; providing an overview vision of the organisation for central managers. But major potentials of big data – analysis to help optimise operations, application of data mining techniques to identify otherwise unseen patterns, use of machine learning techniques to improve strategic decision-making – remain unexploited.

Big data's velocity functionality has delivered faster depot management functions in BMTC, and faster fault rectification in Stelcorp. But other activities have not yet changed: for example, billing and calculation of losses in Stelcorp still happens with a one- or two-month time lag. In Stelcorp the greater variety of data now available is being used to make better planning decisions around losses in the network. But – mirroring the point about data mining and machine learning – the opportunities to extract value by integrating the different data now available have not yet been realised. We can finally interrogate a fourth V – visibility – associated more with open than big data. These corporations could open up their big data, allowing local enterprises to extract more business value from it and becoming more transparent to citizens and civil society organisations. But they have so far chosen not to do this.

Turning to ***outcomes***, there is a vibrant debate in India about data rights; especially in relation to Aadhaar, the unique ID number allocated to all citizens onto which a host of public and private services are being layered. Aadhaar-related arguments are vocal and ongoing relating to issues of data privacy, security, surveillance, access, representation, etc. (Taylor & Broeders 2015, Bhatia & Bhabha 2017). But, possibly because Aadhaar sucks the oxygen away from debates about other big data applications in India, there have been no such arguments relating the Stelcorp and BMTC. The only time these issues were mentioned by interviewees were as rationales for not opening up corporate big data to the public (see also Menezes 2018).

There was evidence, though, of the "shifting terrain of power" associated with the advent of these big data systems. As already suggested in the literature, this included evidence of a "privatisation" of development as relative power moved from public to private sector stakeholders; that is, from the corporations to their ICT partners. This related to changing resource ownership; not of the big data itself but of the surrounding knowledge about big data systems and analytics. Shifts in control over resources were also seen within organisations alongside shifts in other aspects of the Figure 2 model: knowledge and discourse, formal and informal institutional forces, structural position. These underpinned a relative weakening of labour and of middle management in favour of central management that was epitomised by the Panopticon spectacle of control rooms with walls of large display screens visualising the results of real-time big data analysis.

While the realities of control might not be as great as the image, it was the new imaginary of the corporations and of their capabilities that had seeped into the minds of senior managers and politicians. The image and discourse of big data had triggered a growth in "solutionism": the idea of technical solutions to often-complex problems (Taylor & Broeders 2015). For BMTC, this was the problem of its large-scale financial losses. While these had systemic and contextual causes, managers saw the big data systems could help by halving crew numbers on each bus. In Stelcorp, this fed a much wider concern: that of local politicians to be re-elected. While big data in practice had not particularly been associated with a reduction in blackouts and brownouts, the new imaginary it fostered had helped drive a massive expansion in rural electrification.

Summary and Implications

In summary, datafication including use of big data is an increasing reality in developing countries, as these two case studies illustrate. They have been much slower to implement than anticipated, and indicate the scale of implementation challenge facing big data systems in developing countries. The systems are already delivering value for public service organisations but, as yet, more operational than strategic and incremental not transformative. The flows of value are differentiated. They are helping to change profiles of power within and between organisations. And they are intersecting with politics especially around the presented image of organisations and the imaginaries of wider stakeholders.

While care must naturally be taken in generalising from these cases, it is reasonable to conclude that managers and other practitioners must prepare for longer-than-expected project timescales. Intended value may take several years to emerge given the size and complexity of big data systems, and the need to address ongoing socio-technical challenges. The information value chain is useful in mapping the progression of these systems from upstream to downstream focus, including shifts in the focus for key implementation issues. Likewise the pyramidal model provides a simple but helpful insight into the value derived from big data including – one assumes – a slow trajectory from impacting lower-value operational to higher-value strategic decisions. This can only be assumed because of the lack of evidence so far of strategic, transformative impact except though the changed perceptions of senior managers and politicians.

That perceptual change, alongside the evidence about flows of power within and outside the organisations, suggests that Figure 2's model can help understand the relationship between power and big data systems. However, more work would be required to apply the model in greater depth. The same is true of the information value chain and decision pyramid: there is a research agenda to apply all of these to a wider variety of big data systems.

In terms of broader recommendations, there is a need to further develop big data-related capabilities in the public sector. There is a practical aspect to this: a number of challenges might have been avoided or mitigated with greater ability of public officials and technical staff to engage with implementation. As a simple example, lessons learned from these implementations could readily be shared with equivalent corporations in other Indian states and beyond. There is also a political aspect: without greater in-house capacity, the power shift from public to private sectors will remain entrenched, making public organisations dependent on largely unaccountable and unremovable private firms.

Though barely mentioned as yet, these big data systems and the public-private partnerships on which they are built, will need to be more-closely scrutinised around issues of data security, re-use, privacy and surveillance. The managers using these as reasons to resist opening up their data are not entirely wrong, and more effective evaluations – such as data privacy impact assessments – and policies and regulations will need to be put in place. Once that happens, there should be no reason to keep blocking the opening of data: this is data gathered about public processes and assets, paid for by public money, and the economic and political value of data for enterprise and transparency should be maximised.

References

Akshatha, M. (2016) All you need to know about BMTC's Intelligent Transport System, *Citizen Matters*, 26 May http://bengaluru.citizenmatters.in/intelligent-transport-system-bmtc-bangalore-8366

Albanna, B. & Heeks, R. (2018) Positive deviance, big data, and development: a systematic literature review, *Electronic Journal of Information Systems in Developing Countries*, e12063

Ali, A., Qadir, J., ur Rasool, R., Sathiaseelan, A., Zwitter, A. & Crowcroft, J. (2016) Big data for development: applications and techniques, *Big Data Analytics*, 1(2), 1-24

Bhatia, A. & Bhabha, J. (2017) India's Aadhaar scheme and the promise of inclusive social protection, *Oxford Development Studies*, 45(1), 64-79

Bhattacharya, A. (2016a) BMTC to go intelligent, will introduce vehicle tracking, electronic ticketing & passenger information systems, *The Economic Times*, 2 May http://economictimes.indiatimes.com/news/politics-and-nation/bmtc-to-go-intelligent-will-introduce-vehicle-tracking-electronic-ticketing-passenger-information-systems/articleshow/52069334.cms

Bhattacharya, A. (2016b) BMTC flags off intelligent transport system in low-key manner, *The Economic Times*, 25 May http://tech.economictimes.indiatimes.com/news/technology/bmtc-flags-off-intelligent-transport-system-in-low-key-manner/52435841

Blumenstock, J.E., Cadamuro, G. & On, R. (2015) Predicting poverty and wealth from mobile phone metadata, *Science*, 350(6264), 1073-1076

BM (2016) Now, real-time monitoring of BMTC buses for compliance, *Bangalore Mirror*, 26 Mar https://bangaloremirror.indiatimes.com/bangalore/others/Now-real-time-monitoring-of-BMTC-buses-for-compliance/articleshow/51553956.cms

BMTC (2013) *BMTC Gallery*, BMTC, Bengaluru https://mybmtc.com/en/gallery

BMTC (2017) BMTC's Vehicle tracking system consists of Vehicle Tracking Unit & Voice Kit for locating the bus & communicating with the driver respectively, *Twitter* 16 Mar https://twitter.com/BMTC_BENGALURU/status/842619833938329600

BMTC (2018a) *BMTC at a Glance*, BMTC, Bengaluru http://www.mybmtc.com/en/bmtc_glance

BMTC (2018b) *Performance Indicator*, BMTC, Bengaluru http://www.mybmtc.com/en/performance-indicator

Boellstorff, T. (2013) Making big data, in theory, *First Monday*, 18(10), 1-17

Boyd, D. & Crawford, K. (2012) Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon, *Information, Communication & Society*, 15(5), 662-679

CM (2017) One year later, BMTC's updated ITS continues to disappoint, *Citizen Matters*, 24 May http://bengaluru.citizenmatters.in/one-year-later-bmtc-intelligent-transport-system-continues-to-be-unintelligent-9308

Dumbill, E. (2012) What is big data?, *O'Reilly*, 11 Jan

Heeks, R. (2017) *A Structural Model and Manifesto for Data Justice for International Development*, 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: what would it mean?, *Information Development*, 34(1), 90-102

Hickey, S. & Mohan, G. (2005) Relocating participation within a radical politics of development, *Development and Change*, 36(2), 237-262

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

Jha, M. (2017) Big data, real-time GPS: tonic, not steroids, for Bengaluru's bus service, *Factor Daily*, 27 Oct https://factordaily.com/longform/big-data-real-time-gps-tonic-not-steroids-bengalurus-bus-service/

Kidiyoor, S. (2018) BMTC takes a big hit as ridership falls, *The Hindu*, 3 Jul https://www.thehindu.com/news/cities/bangalore/bmtc-takes-a-big-hit-as-ridership-falls/article24317525.ece

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

Lokanathan, S. & Gunaratne, R.L. (2015) Mobile network big data for development: demystifying the uses and challenges, *DigiWorld Economic Journal*, 97(1), 75-94

Menezes, N. (2018) How BMTC has failed to live up to its tech expectations, *Economic Times*, 14 Dec https://tech.economictimes.indiatimes.com/news/technology/how-bmtc-has-failed-to-live-up-to-its-tech-expectations/67086053

Min, B. & Golden, M. (2014) Electoral cycles in electricity losses in India, *Energy Policy*, 65, 619-625

Mishra, B. (2016) Intelligent Transport System (ITS), presentation at workshop on *Smart Mobility for Bengaluru*, Bengaluru, 10 Jun https://www.slideshare.net/EMBARQNetwork/bmtc-intelligent-transport-system

Murphy, J.T. & Carmody, P. (2015) *Africa's Information Revolution*, John Wiley, Chichester, UK

Narasimhan, S.R., Seshadri, P., Raoot, M.G. & Khaparde, S.A. (2011) Indian electricity grids: a decade of reformation and roadmap ahead. In: *2011 IEEE Power and Energy Society General Meeting*, IEEE, New York, NY, 1-6

Philip, C.M. (2017) BMTC plans to trim manpower with conductor-less buses, *Indian Express*, 13 Feb http://www.newindianexpress.com/cities/bengaluru/2017/feb/13/bmtc-plans-to-trim-manpower-with-conductor-less-buses-1570054.html

Rakesh, V., Heeks, R., Chattapadhyay, S. & Foster, C. (2018) *Big Data and Urban Transportation in India: A Bengaluru Bus Company Case Study*, GDI Development Informatics Working Paper no.72, University of Manchester, UK

Ramakrishnan, S. (2016) A sneak peak at BMTC's and MCTD's ITS, *BESTpedia*, 18 Jun http://bestpedia.in/2016/06/sneak-peak-bmtc-mctd-its/

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

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

Turban, E., Pollard, C. & Wood, G (2018) *Information Technology for Management*, 11th edn, John Wiley & Sons, Hoboken, NJ

UNGP (2014) *Mining Indonesian Tweets to Understand Food Price Crises*, UN Global Pulse, Jakarta

World Bank (2014) *Big Data in Action for Development*, World Bank, Washington, DC