Exploring Big Data for Development: An Electricity Sector Case Study from India

Paper Category: Research-in-progress

Ritam Sengupta
Centre for Studies in Social Sciences, Calcutta, India

Richard Heeks[1]
Centre for Development Informatics, University of Manchester, UK

Sumandro Chattapadhyay
Centre for Internet and Society, Bengaluru, India

Christopher Foster
Information School, University of Sheffield, UK

ABSTRACT

This paper presents exploratory research into "data-intensive development" that seeks to inductively identify issues and conceptual frameworks of relevance to big data in developing countries. It presents a case study of big data innovations in "Stelcorp"; a state electricity corporation in India. In an attempt to address losses in electricity distribution, Stelcorp has introduced new digital meters throughout the distribution network to capture big data, and organisation-wide information systems that store and process and disseminate big data.

Emergent issues are identified across three domains: implementation, value and outcome. Implementation of big data has worked relatively well but technical and human challenges remain. The advent of big data has enabled some – albeit constrained – value addition in all areas of organisational operation: customer billing, fault and loss detection, performance measurement, and planning. Yet US$ tens of millions of investment in big data has brought no aggregate improvement in distribution losses or revenue collection. This can be explained by the wider outcome, with big data faltering in the face of external politics; in this case the electoral politics of electrification. Alongside this reproduction of power, the paper also reflects on the way in which big data has enabled shifts in the locus of power: from public to private sector; from labour to management; and from lower to higher levels of management.

A number of conceptual frameworks emerge as having analytical power in studying big data and global development. The information value chain model helps track both implementation and value-creation of big data projects. The design-reality gap model can be used to analyse the nature and extent of barriers facing big data projects in developing countries. And models of power – resource dependency, epistemic models, and wider frameworks – are all shown as helping understand the politics of big data.

Keywords: big data; developing countries; political impact; ICTs and infrastructure; electricity; India

[1] Corresponding author: richard.heeks@manchester.ac.uk

INTRODUCTION

We are seeing emergence of "data-intensive development": an ever-growing role within global development for data generally and for particular forms such as big data, open data and real-time data (IEAG 2014, Heeks 2015). But the hype on data-intensive development is running ahead of practice; and practice in turn is running ahead of research (Heeks 2016). This therefore creates a number of knowledge gaps that need to be filled if we are to understand how data can fulfil its potential for global development.

To help address this, in 2016, the Universities of Manchester and Sheffield collaborated with the Centre for Internet and Society in India to commission a small set of pilot studies. Their specific focus was big data within three sectors that already had initiatives underway: electricity, education, and transport. The aim of the studies and, hence, the aim of this paper (which reports on the electricity study) was to identify the issues that emerge when big data is implemented in a developing country context; and to identify conceptual frameworks that are relevant for analysis of data-intensive development; in particular, big data initiatives.

The next section provides a general overview of themes within the literature on data-intensive development; specifically, that dealing with big data. Then, the case study methodology for this exploratory, inductive research is outlined. Case study findings are presented before final discussion and conclusions that link back to the main aims around emergent issues and conceptual frameworks.

LITERATURE REVIEW

The well-known characterisation of big data sees it as different from traditional data in terms of three Vs, though only one of these is actually related to its bigness: volume (a greatly-increased amount of data available), velocity (greatly-increased speed with which data is available), and variety (greatly-increased number of forms and types of data) (Kshetri 2014).

In looking at the relatively-nascent literature on big data and development, we can categorise it into four main domains, as summarised in Figure 1 (Heeks & Sampaio 2016). These are developed on the basis of the "information value chain": the steps required to turn data into an impact on global development.

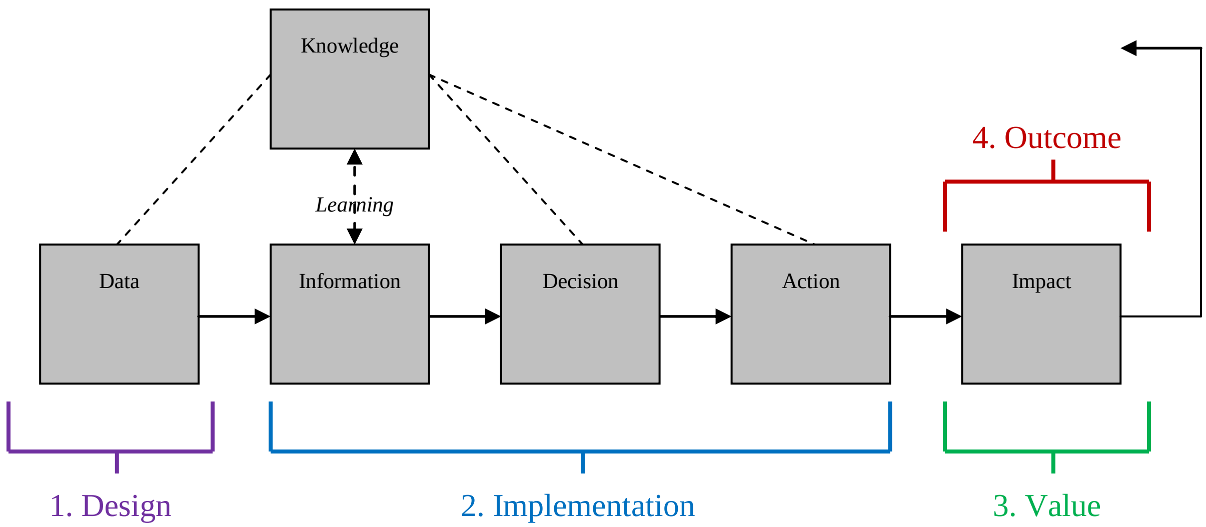


Figure 1: Domains of Literature on Big Data and Development

Design-related research looks at the particular issues that arise when seeking to design big data initiatives for global development purposes. Work to date has tended to focus on specific technical issues of design – for example, around cleaning and standardising big data sets (e.g. Blumenstock et al 2016), or around development of data analytical techniques (e.g. Jahani et al 2015). While there is recognition that there may be some greater challenges here in developing countries, there does not yet appear to have been much consideration or conceptualisation of the specific context of developing countries or the specific purposes of global development. There is clearly potential for this to occur given the work within the wider ICT4D domain on design methodologies for development (e.g. Dearden & Rizvi 2015).

Implementation-related research moves beyond the purely technical to investigate how big data can be put into practice within development settings: turning the data into information that can be used for decisions which are operationalised into actions. There are still specific technical issues considered in the literature such as visualisation of big data for development decision-makers (e.g. Isafiade & Bagula 2013). But the leitmotif of literature in this domain is that of gaps and barriers: shortfalls in required capacities that are preventing effective implementation of big data initiatives. For instance, Hilbert (2016) speaks of a "big data divide" with technical, services and human dimensions. Gomez & Heeks (2016) identify Hilbert's three-dimensional categorisation as the only analytical frame dealing with implementation barriers to big data and development, though they identify potential value in Heeks' design-reality gap model; showing how it can provide a broader foundation for analysis.

Value-related research investigates the organisational and developmental value purposively created by big data. Although there is a range of literature dealing with this, it largely falls into two camps. It is either specific proof-of-concept research that shows the potential for a contribution; for example, demonstrating the feasibility of using social media to track food prices (PLJ 2014). Or it is 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. World Bank 2014, Hilbert 2016). The conceptualisation of value tends to be simplistic if present at all: for example, categorising value by development sector or by type of decision.

Outcome-related research looks at the broader implications of big data on development and differs from the focus of value in the way that "immanent development" – emergent and underlying – differs from "imminent development": purposive and intentional (Hickey & Mohan 2005). The corpus of work is small and approaches this domain from different directions including rights (McDonald 2016), justice (Heeks & Renken 2016) and politics (Taylor & Broeders 2015). Given the tendency of most practice to be only just moving through the design and implementation phases, there is little direct evidence available. Most work at present is therefore based more on extrapolation and identification of likely issues; especially framed around inequality such as loss of rights or loss of power by particular groups, or around shifts in power and rights between groups in developing countries. There are a few proposed conceptualisations, such as those around the epistemic power of big data: its ability to change representations and understandings of real-world systems (*ibid.*).

The review confirmed the formative nature of the research field, and the need for much more research dealing with actual case studies of big data, such as that reported here. As explained next, because of the formative nature of both the research field and our case study knowledge, we took forward from the literature review the overall frame of four domains, and a sensitisation to the particular issues that each domain had raised within the literature. We did not take forward a specific conceptual framework for analysis because of the breadth of scope of our case exploration: beyond the range of any individual framework. Instead, we were seeking to understand what types of framework would be appropriate to analyse different aspects of the case study.

METHODOLOGY

Given the formative nature of both research and practice in relation to big data and development, we decided it would be best to adopt a case study research strategy. Case studies are especially fitting for research such as ours where the researcher has "less a priori knowledge of what the variables of interest are" and wishes to develop an initial understanding that builds upwards from practice towards conceptual framing (Benbasat et al 2002:81). We identified our case studies iteratively: consulting our contact network in India to find examples of organisations that were said to be applying big data, and conducting an initial contact interview with an organisational representative to ascertain if this was correct.

The case study reported here relates to a big data initiative in a large Indian state-owned power corporation – anonymised as "Stelcorp" and based in "Janakari" State – which serves well over 15 million customers (c.80% in rural areas), employs more than 15,000 staff, and has revenues in excess of US$1bn per year. Several of the technical innovations have been public-private partnerships between Stelcorp and a number of different Indian-owned, private sector IT firms. In particular, Stelcorp has worked with a large IT sector corporation, "Digicorp".

From this point of case study selection, we adopted a largely inductive approach. As noted above, we had been sensitised by literature review to potential domains and issues. However, our lack of prior contact with the case study organisations and lack of any research literature on these cases meant we could not predetermine the particular domains and issues that the case would generate. Nor could we therefore adopt a deductive approach with a priori selection of a specific theory or model to apply; even setting aside the relative lack to date of theory-based work on big data and development that currently hampers deductive work in this field. Over the longer-term of engagement with our case organisations, we intend to adopt an abductive approach, but here we report only the inductive first stage seeking – as per the stated aim of the paper – to identify emergent issues and conceptual frameworks of relevance.

Given the stage of our investigation, and consistent with the tenets of case study research (Yin 2012), we designed an approach that would be a mix of the descriptive and the exploratory; relating to different domains of big data and development. The descriptive component would aim to ask the "What" and "When" of the big data innovations, to produce a design and implementation chronology. The exploratory component would be more open and seek to understand the value and outcome issues arising within the case study.

Again consistent with case study research, we adopted a multi-method qualitative approach that attempted some triangulation (*ibid.*). We undertook eight interviews with Stelcorp senior and mid-level managers and engineers responsible for strategy, finance, planning and information systems. These interviews used a mix of semi-structured elements to cover the descriptive components described above, and more unstructured discussion to cover the exploratory components.

Private partners, politicians and fieldworkers were not available for interview and, as a result of lack of access to Digicorp, we were not able to develop findings related to the design domain. However, the views of all these groups were represented through analysis of documentation from these sources or their representatives. There was therefore some stakeholder triangulation (though views of consumers were not directly reflected). Methodological triangulation was made possible through some limited observation of activity in Stelcorp facilities and through use of documentary sources including:

- Stelcorp annual reports, strategic plans and tender documents,
- documentation from audit and regulatory authorities at state-level,
- overview reports from the central Ministry of Power,
- correspondence and newsletters made available from trade unions and staff associations within Stelcorp,
- published case studies and summaries from private partners, and
- state-level electricity policies.

The requirement for case anonymity has meant it is not possible to cite some of these documents since to do so would enable case identification.

To analyse these data sources, and reflecting the developmental stage of research, we did not formally code but, instead, categorised material into the three active domains – implementation, value, and outcome – and iteratively discussed and developed emergent issues among the research team.

CASE STUDY FINDINGS

As discussed above, emergent issues and potential conceptualisations are discussed in relation to the three field domains: implementation, value and outcome. First, though, we step back to understand the nature of the wider context into which this big data initiative was being introduced.

Context: The Electricity Sector in India

Power – particularly electrical power – is an essential component of development (Jacobson et al 2005). Yet both expansion and operation of electricity supply in developing countries have been plagued by problems; problems that have had a direct constraining effect on both economic and social development (Winkler et al 2011). India is no exception. It has struggled to expand generating capacity, it has struggled to effectively distribute the electricity that it generates to users, and it has struggled to recoup costs from the users it does serve (Srivastava & Rehman 2006).

An average one-third of power generated in India is lost; "estimated to cost India's economy 1.5% of GDP each year, aggravating chronic power shortages and straining the precarious finances of its public electricity providers" (Min & Golden 2014:619). "Aggregate technical and commercial (AT&C) losses", as they are known, are calculated as the percentage difference between the power injected into a distribution grid and the revenue collection realised in equivalent units. 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; a rather polite term for what is typically a mix of incompetence and theft including "inefficiencies in billing, meter tampering, illegal connections, and use by flat rate consumers that exceeds their nominal allotments" (*ibid.*).

Recognising the scale of the problems and of the needed solutions, India has been steadily reforming its power sector since the early 1990s with a combination of more political and more technical reforms as reflected, for example, in the 2005 National Electricity Policy (Bajaj & Sharma 2006). The main thrust of the policy reforms has been neo-liberal: attracting private finance into the sector; allowing private sector power generation; unbundling of generation, transmission and distribution into separate entities; privatisation; etc. In addition, there have been more technical reforms such as the 2001 Accelerated Power Development & Reform Programme which was revised in 2008 as the Restructured Accelerated Power Development and Reforms Programme (R-APDRP) (Narasimhan et al 2011).

R-APDRP is a two-part initiative aimed at reducing AT&C losses to 15% by 2020, mainly focusing on urban and peri-urban areas (MoP n.d., Arthapedia 2015). Part A is fully funded by central government and involves the installation of digital information systems to help with both technical and commercial aspects of loss: power monitoring systems, energy accounting systems, billing systems, customer grievance redress systems, etc. Part B is only part-funded and aims to renovate and upgrade the distribution networks, for example to help balance loads between parts of the network or to improve cabling to reduce losses.

Technical losses through heat/resistance are inversely proportional to voltage: hence the use of high voltages for transmission over long distances, with voltage only being stepped-down close to the point of use (see Figure 2). This means most heat-related technical losses occur at the lower levels of the distribution network. The same is true of fault-related losses given the proliferation of technical components at these lower levels: as an example, a typical failure rate of transformers in India was reported to be 6% over a six-month period. Likewise commercial losses all occur at the lowest level of the network among the millions of consumers. As a result, R-APDRP activities also deal with that part of the network and are in consequence very large in scope and number and geographic dispersion. In further consequence, Part A of this scheme therefore represents the arrival of big data into Indian electricity distribution; reflecting the three "Vs" of big data by capturing a far greater amount of data from large numbers of sources about several different entities and delivering it at much greater speed to the organisation for subsequent processing and use.

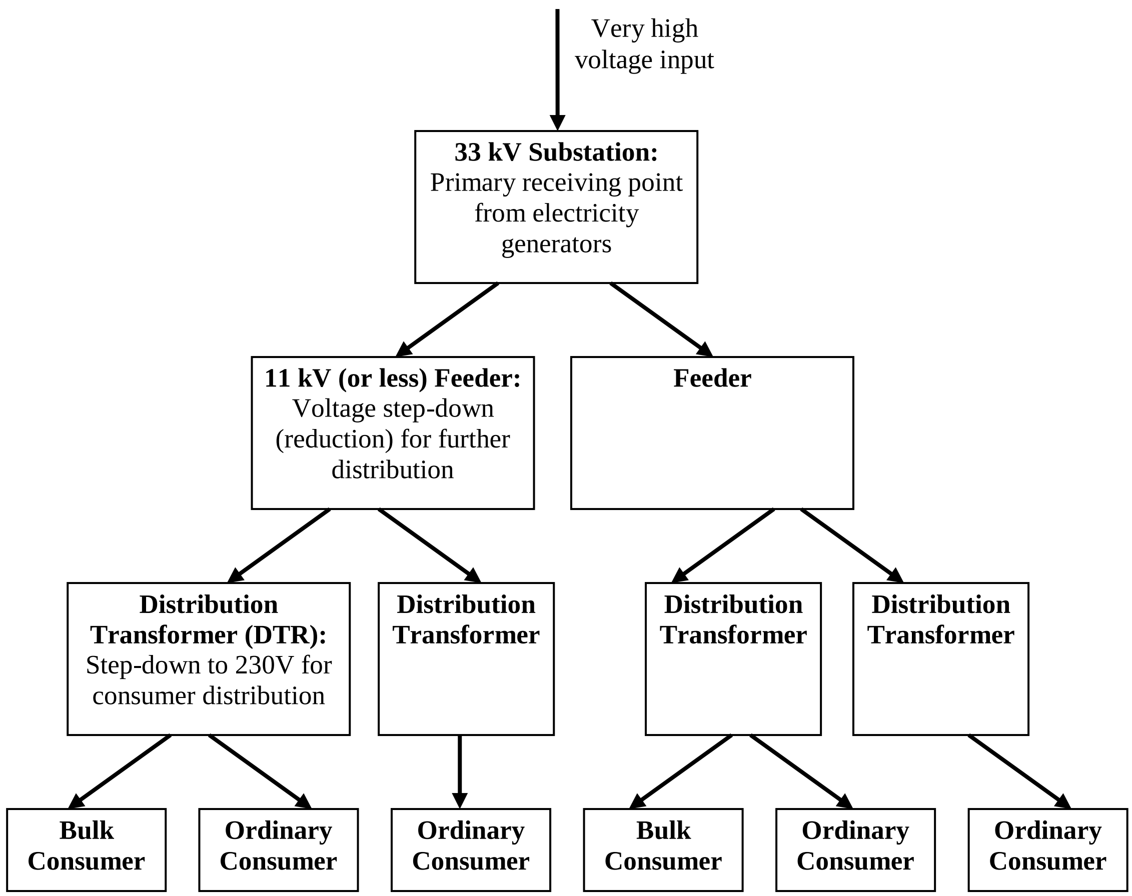


Figure 2: Schematic Electricity Distribution Network

Implementation: Data-Related Innovations

The data-related technical innovations implemented by Stelcorp in recent years are described below, with a summary timeline presented in the Appendix.

Data-Capture Innovation

During the 20^{th} century, most data about the distribution network was captured manually; for example by having a human meter-reader take down on paper the details from analogue meters that existed within the network (see Figure 3; Pradesh18 2015).

Figure 3: Paper-Based Recording of Analogue Meter Data

From the turn of the century, Stelcorp began a three-element programme of installing new digital meters. For the higher levels of the power network – and particularly since 2009 and R-APDRP – they installed automated, online digital meters. These are installed at all urban feeders (which typically cover a population of around 25,000; see Figure 2), for all urban distribution transformers (which typically cover 50-100 consumers), and for all bulk power consumers (such as factories, food processing plants, cold storage facilities, etc). At the time of writing, more than 3,000 feeders and nearly 200,000 DTRs (including rural: see next) are said to be metered.

These meters (see Figure 4; IndiaMart n.d.a) can capture and transmit a variety of network "health parameters" such as regular measurements of voltage, current, load and energy; event alerts such as tampering or power outages; and can log data over a 45-day period to provide a historical picture e.g. of feeder/transformer load over time. They are connected via wireless GSM modems and push data at regular intervals – 15-minutely for feeders, daily for other installations – but can be accessed in real-time as communication is bi-directional.

There have been some problems in practice. Sourcing meters from different producers has made it difficult to standardise data into a common readable format: some still has to be re-assembled by human intervention. And Stelcorp engineers complain about unreliability of meter connectivity and operation (claimed by some to affect as many as half the meters) including provision of impossible data readings by some meters (e.g. showing actual power carried higher than the theoretically-possible maximum). One report claimed that, while 100% of feeder metering was working, this was true of only 30% of transformers.

Figure 4: Typical Feeder/DTR Meters

Though not part of R-APDRP, on its own initiative Stelcorp also from 2009 began installing stand-alone meters (not readable remotely) on all rural feeders and transformers. These are the same type of meter as shown in Figure 4, except they do not have GSM connectivity. They therefore have to be read manually by human meter-readers who, in theory on a monthly basis, offload data from the meter onto a "Common Meter Reading Instrument" (CMRI: similar to that shown in Figure 5; IndiaMart n.d.b) and then upload it at their divisional centre.

Figure 5: Common Meter Reading Instrument

Finally, at the level of ordinary consumers Stelcorp has since the mid-2000s been replacing analogue with digital meters that are relatively more tamper-proof. At the time of writing, virtually all consumer meters were now of this type, though they are still fairly basic in just recording the level of electricity use. There has also been a change in data capture: as with the rural meters described above, data from consumer meters is now captured by human meter-readers offloading that data onto CMRIs. The data is then uploaded at local centres largely for billing purposes: most customers are billed quarterly on the basis of their electricity consumption. Meter-readers should also record data on consumers using electricity without meters or without being listed on the billing system, on defective meters, on customers who have moved, etc. In theory, digital recording of consumer meters was very close to completion. However, there were indications that at least some of this data was still being recorded on paper and that, rather than going out to capture data direct and reading meters on a monthly basis as intended, some meter-readers were just logging average values based on consumption levels in earlier months.

Data-Systems Innovation

Stelcorp operates a four-level organisational hierarchy starting with central headquarters in the Janakari State capital. The state is then split into around half-a-dozen zones which in turn give way to dozens of divisions, and hundreds of local centres. Until the 1990s, data input and communication was largely paper-based and localised. Meter-readers would write down data from analogue meters and record them in a master register at their local centre; some limited aggregate data would be passed up the organisational hierarchy. From the late 1990s, this data was input into newly-installed computer systems at divisional level. From the mid-2000s, this was passed up to newly-installed computer systems at zonal level. In the 2010s, a new central data centre was created and linked to zones via mid-speed broadband internet connections and to districts and local centres via low-speed broadband. In order to cope with the quantities of data being produced by the new data-capture technologies, the central servers have capacities building upwards from hundreds of terabytes.

This organisation-wide digital network has allowed development of the Meter Data Acquisition System (MDAS) by private sector IT partner – "Digicorp" – appointed in 2009, with MDAS functional since 2012. Digicorp is responsible for designing and operating the software architecture underpinning MDAS, for data management, and for training of Stelcorp staff. Data from the online meters at the higher levels of the power network is input automatically into MDAS. Data from the other digital meters is input each month into the meter-readers CMRIs, and then directly uploaded into the data system at a local or divisional centre, with an overall summary as shown in Figure 6. MDAS therefore acts as the organisational data repository, storing 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.

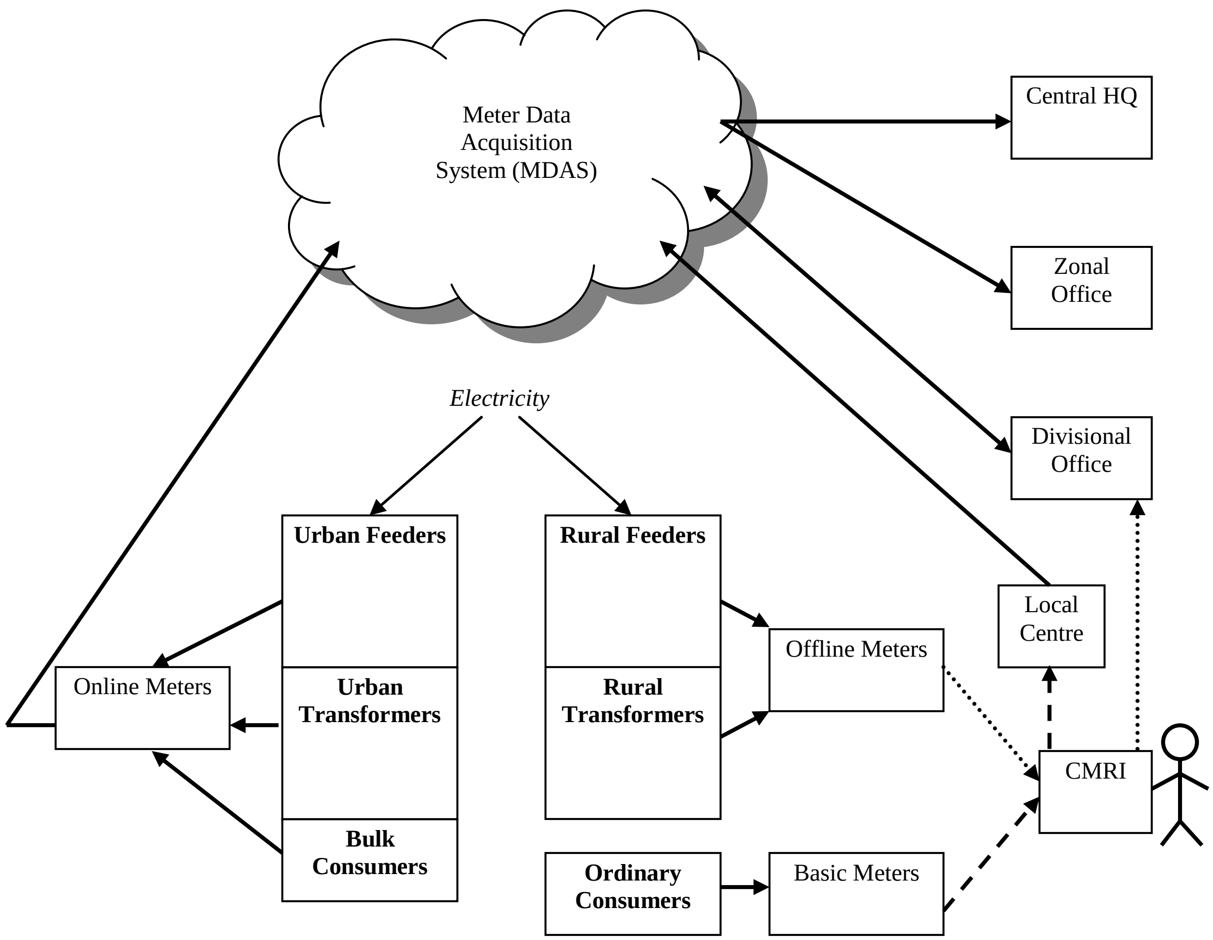


Figure 6: Stelcorp's Data System

With MDAS as the data foundation for Stelcorp, it then becomes possible to layer other information systems on top. For example, in 2012, Digicorp implemented a consumer billing information system based around IS-U: Germany multinational SAP's enterprise resource planning system specifically designed for utility companies. However, as reported by Stelcorp staff, this ERP application has run into difficulties – including consumers remaining unbilled – because it was designed on the basis of a common meter reading and billing cycle for all consumers linked to a specific DTR; i.e. everyone billed on the same day each quarter. However, both practice and the prior data system within Stelcorp were based around different meter reading and billing dates for all consumers with no common cycle for those connected to one transformer. Given consumer-level data is also used for calculating transformer-level losses, this may also have a negative knock-on to the accuracy and availability of this type of loss data.

Data-Output Innovation

In collaboration with a private sector partner, since 2010 Stelcorp has been geo-locating all of its more than four million assets (feeders, DTRs, pylons, meters, etc) alongside since 2014, tracing and developing a unique ID code for every individual customer's connection from the local transformer. This allows map-based display of all assets and events using a geographic information system. Although declared completed in 2015, interviewees stated that some errors in both geo-location and customer ID coding regularly emerge – for example with new lines and new DTRs not being incorporated – and this therefore appears to be something of a work in progress. A report in 2016 said 90% of consumer-linked transformers had been mapped.

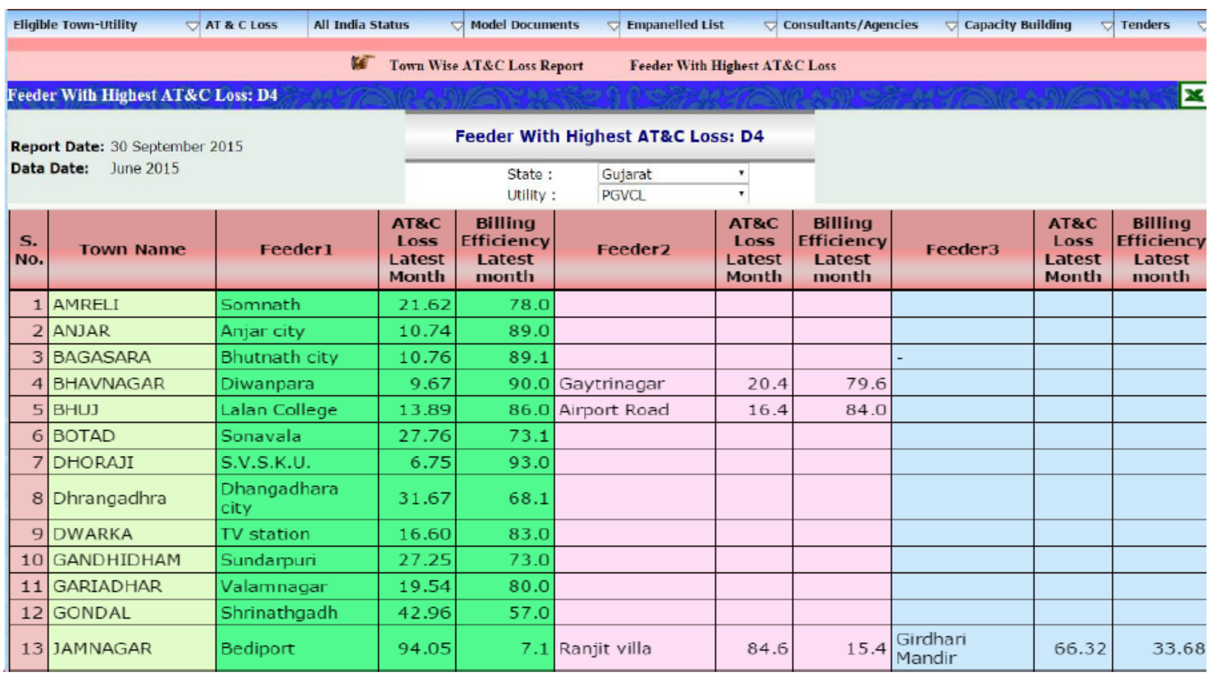
Implementation: Challenges and Frameworks

The innovations described above provide a big data system foundation for Stelcorp. This foundation is not without its problems with speed and accuracy continuing to be compromised by issues with meter reading errors/unreliability, with GIS/ID coding errors, and with the clash over meter reading cycles between billing system design and the actual practices of Stelcorp. All this has created an ongoing requirement for human intervention in parts of the system that are intended to be fully automated. While all these issues are likely to be transitional rather than permanent, they not ephemeral: instead, they appear inherent in moving to big data-based operations.

In looking for potential conceptual frameworks to help understand these challenges, we identified one option as the design-reality gap model (see Gomez & Heeks 2016). This could be used to interpret challenges as a gap between the design assumptions and requirements of big data as compared to the actual realities of organisations like Stelcorp. For example, design requirements for reliability of technical infrastructure that were not quite met in reality, or design requirements for accuracy of existing data that was not quite met in reality. This was also seen with the enterprise resource planning system, and design-reality gaps are well-known to afflict ERP applications in developing countries (e.g. Hawari & Heeks 2010). Such applications seek to impose a standardised view of the world: a view inscribed into the systems from the private sector of the global North that can easily mismatch the localised realities of organisations in the global South. The example described above is illustrative: a design based on common billing cycles for all customers on one transformer vs. a reality of individualised billing cycles for all customers.

Value: The Impact of Big Data-Supported Decisions and Actions

Overall, the data being collated into MDAS can be viewed as a "dashboard" showing the various data items that are now being collected – power health parameters like current and loading, AT&C losses, billing and revenue – displayed by slicing and aggregating / disaggregating in various different ways. An example is provided in Figure 7, which shows AT&C loss levels for feeders in particular towns (MoP 2015a)[2]. The example shown is typical of the way in which MDAS is used: as a management information system delivering routine overviews of aggregated, time-lagged data.



S. No.	Town Name	Feeder1	AT&C Loss Latest Month	Billing Efficiency Latest month	Feeder2	AT&C Loss Latest Month	Billing Efficiency Latest month	Feeder3	AT&C Loss Latest Month	Billing Efficiency Latest month
1	AMRELI	Somnath	21.62	78.0						
2	ANJAR	Anjar city	10.74	89.0						
3	BAGASARA	Bhutnath city	10.76	89.1				-		
4	BHAVNAGAR	Diwanpara	9.67	90.0	Gaytrinagar	20.4	79.6			
5	BHUJ	Lalan College	13.89	86.0	Airport Road	16.4	84.0			
6	BOTAD	Sonavala	27.76	73.1						
7	DHORAJI	S.V.S.K.U.	6.75	93.0						
8	Dhrangadhra	Dhangadhara city	31.67	68.1						
9	DWARKA	TV station	16.60	83.0						
10	GANDHIDHAM	Sundarpuri	27.25	73.0						
11	GARIADHAR	Valamnagar	19.54	80.0						
12	GONDAL	Shrinathgadh	42.96	57.0						
13	JAMNAGAR	Bediport	94.05	7.1	Ranjit villa	84.6	15.4	Girdhari Mandir	66.32	33.68

Figure 7: Example of MDAS Interface Showing Feeder AT&C Loss Measurements

[2] This is a generic example from Gujarat provided via the Ministry of Power's website, and not from Stelcorp's system.

Billing

The volume and velocity functionalities of big data provide Stelcorp the potential for faster and more accurate billing of consumers. "Spot billing" has now become feasible, whereby field staff read meters and issue an accurate bill on the spot to customers (though, despite the potential, they do not seem to then collect payment on the spot). This has now become part of the billing portfolio of Stelcorp (alongside regular billing) at almost all of its local centres though still not as widespread as intended. In general, regular billing is more accurate than previously, though with the caveats around data quality from customer meters noted above, but it is necessarily not faster since it occurs with the same i.e. quarterly, cycle times as before. This reflects the generally limited impact of velocity functionality: other than use of online meter data for fault rectification (see next), much of the upload and calculation activity relating to MDAS in terms of billing or calculation of losses happens with a one- or two-month time lag.

Fault and Loss Detection and Rectification

The velocity, volume and variety functionalities of big data have enabled some innovation relating to fault and loss detection. These are separate but related issues: faults are very short-term power outages caused by equipment malfunction; losses are identified on a longer-term basis by retrospective calculation and may have multiple causes which could be technical or human. Previously, faults leading to power outages would be reported either by customers or by staff in the field, and faults or problems leading to longer-term AT&C losses would be identified through the results of manual calculations. Detection of the fault location in either case would then be a laborious trial-and-error process undertaken by linesmen (and they are all men) testing "health parameters" over a large area – and based on a mix of guesswork, intuition and experience – until the fault location was found and the fault rectified. Initial computerisation involved manual data recording, with losses reported at the level of divisions. This meant that longer-term loss identification became rather less hit and miss but specific fault detection and rectification was still a manual process.

Under MDAS, there is still no automatic, real-time reporting of faults. Instead, outages mainly continue to be reported from outside e.g. with customers calling or texting their local centre, or with linesmen identifying a problem. Likewise, loss identification is not real-time but retrospective, but has become more comprehensive. Following short-term fault or medium-term loss detection, managers can now drill down into MDAS to identify the likely location of the problem, and pick up related data about the functioning of relevant assets like feeders and transformers if these are online (hence, excluding rural areas). Thus alongside data velocity it is data granularity that has been of particular assistance.

For faults, a "docket" would be opened initially at local level and for losses, a docket would be opened directly at zonal level. This would include a note of the GIS-based location, with details sent via SMS to the linesmen in the appropriate mobile maintenance unit/van (see Figure 8; Stelcorp screenshot). As a result linesmen are now able to go fairly directly to the source of the problem, knowing both its location and nature. This should also be possible with detection of theft-related losses: transformers showing losses can be targeted knowing the codes of all consumer meters linked to that transformer, and usage patterns checked for possible irregularities. Again, the particular benefit arises due to the additional granularity being provided by the big data initiatives.

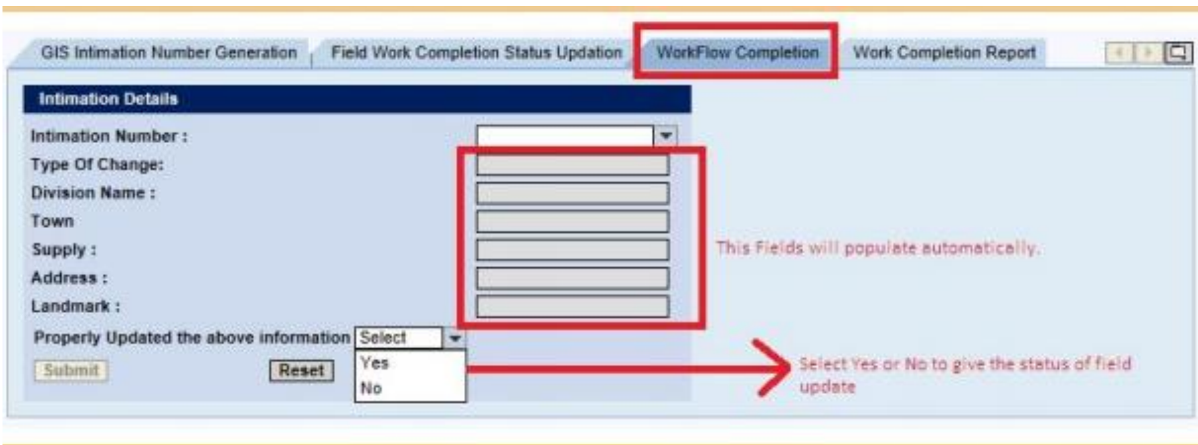


Figure 8: Example of MDAS Interface Showing Form to Trigger Field-Level Work

If the docket is not attended to within a set time period via a work completion report, there is an automated process of escalation by sending emails to higher levels of management including – see below – a central management unit. (The same workflow exists for other forms of customer complaint: they are logged via local centres, a docket is created, and the complaint needs addressing within a particular time period.) Thus, alongside better and faster location of problem sources and transmission to field staff and resolution, there is also better monitoring of the process.

In all, this workflow is therefore quasi-real-time – significantly sped up compared to earlier ways of working – but not yet fully-real-time. The information value chain for this workflow 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.

Performance Measurement

Alongside these data- and workflow-related innovations, there has also 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 the data and workflow innovations described above including: the health parameters of all feeders and transformers in the network; adherence to consumer billing cycles including levels of collection and disconnections; and surveying whether dockets are attended to on time. Alongside this oversight function, FEMU itself performs multi-variate audits that use MDAS and the SAP IS-U application to calculate relative performance tables – both cross-sectional and trends over time – of zones and divisions for energy health parameters, AT&C losses, and billing and collection efficiency[3].

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.

As well as the inwards/downwards role, FEMU also has an upwards/outwards role of reporting Stelcorp aggregate performance data and future targets to regulatory bodies for tariff determination, and to audit and central government funding agencies for financial approval. At central government level, this therefore enables performance league tables of distribution companies to be formulated. An example – for AT&C losses – is shown in Figure 9 (MoP 2015a).

---

[3] Billing efficiency represents the proportion of total energy supplied which is billed to consumers; collection efficiency represents the proportion of the billed amount that is actually collected from consumers (Mam et al 2014).

Rank	States/UT	Name of Utilities	Q1 FY 16
1	AP	APEPDCL	5.3%
2	Karnataka	BESCOM	8.5%
3	Karnataka	CESC	9.0%
4	Karnataka	MESCOM	9.1%
5	AP	APSPDCL	11.4%
6	Telangana	TSNPDCL	12.8%
7	Telangana	TSSPDCL	13.6%
8	Gujarat	MGVCL	13.8%
9	Gujarat	UGVCL	13.8%
10	Kerala	KSEBL	15.9%
11	Goa	ED Goa	16.4%
12	HP	HPSEBL	16.5%
13	Gujarat	DGVCL	17.6%
14	Karnataka	GESCOM	18.4%
15	Karnataka	HESCOM	19.2%
16	Tamil Nadu	TANGEDCO	19.7%
17	Maharashtra	MSEDCL	20.4%
18	Rajasthan	JDVVNL	21.0%
19	Gujarat	PGVCL	23.5%
20	MP	MPPKVVCL	24.8%
21	Meghalaya	MePDCL	26.7%
22	Punjab	PSPCL	26.9%
23	Chattisgarh	CSPDCL	27.0%
24	Haryana	DHBVNL	27.0%

Rank	States/UT	Name of Utilities	Q1 FY 16
25	MP	MPPuKVVCL	28.1%
26	Assam	APDCL	29.1%
27	Rajasthan	AVVNL	29.3%
28	UP	KESCO	31.7%
29	Bihar	NBPDCL	32.7%
30	MP	MPMKVVCL	33.6%
31	Tripura	TSECL	36.1%
32	UP	PVVNL	36.2%
33	Uttarakhand	UPCL	36.3%
34	Rajasthan	JVVNL	36.3%
35	UP	MVVNL	37.0%
36	Sikkim	PD Sikkim	39.5%
37	UP	PUVVNL	40.8%
38	Jharkhand	JhVVNL	41.3%
39	West Bengal	WBSEDCL	42.0%
40	Haryana	UHBVNL	43.3%
41	UP	DVVNL	47.9%
42	Bihar	SBPDCL	49.4%
43	J&K	JKPDD	64.7%
44	Manipur	MSPDCL	70.4%
45	Nagaland	PD Naga	72.3%
46	Mizoram	ED Mizo	-
47	Puduchery	ED Pudu	-
48	Ar Pradesh	PD ArPr	-

Figure 9: National Performance League Table for AT&C Losses

Finally, we should note that under R-APDRP there are some initial pilot installations of SCADA (supervisory control and data acquisition). This should eventually allow for not only remote monitoring of data, as currently (the upstream part of the value chain), but also remote control of some aspects of the electricity network (i.e. making decisions remotely and enacting these as local actions such as re-routing electricity supply: adding in the downstream part of the chain). This is Stelcorp's toe in the water towards the smart grid concept (though a truly smart grid is one that can take some decisions and actions autonomously without human intervention). Foundational components for smart grid are already in place (MoP 2015b): online metering is a prerequisite since that data would be the basis for automated decisions and control over metered assets; and the GIS-based individual coding of consumer connections is a prerequisite for demand-side management since it can be built on to offer consumers algorithmically-generated, price-incentivised options for varying their consumption according to supply-side variations in peak load (in other words, pricing electricity more cheaply at low-load points and more expensively at high-load points). As yet, though, smart grid remains something largely for India's future rather than being any kind of current reality except for very small pilot projects dotted around the country (Selvam et al 2016).

Planning

As noted, FEMU's internal meetings are intended not just to include retrospective review but also prospective short-term planning for corrective actions; what one could see as tactical planning. An example might be planning for actions on a feeder that through its online metering has been showing a high level of technical losses; those actions typically being installation of additional capacitor banks or "bifurcation": the addition of a second feeder to help reduce and balance power loading on the first feeder. In the past, a linesman would have been sent out to the feeder to take thrice-daily measurements and multiply by pre-designated voltage standards to produce an overall gross statistic on which planning was based. This typically resulted in inaccurate (generally under-) estimates of required load, leading to partly-ineffective corrective actions.

Under MDAS, data for any individual feeder can be pulled from the system and will show all the required data across many days of recording. Instead of making decisions on single overall values, they can now be made in consideration of maxima, minima and patterns of load, current and voltage fluctuation. Decisions have thus moved to a richer, multi-variate basis, reflecting the variety functionality of big data. This allows more accurate planning of the particular corrective actions to be taken, though the decision making is still undertaken by humans. Decision making is also faster than previously thanks to ready access to the velocity functionality of this big data.

MDAS also feeds into more strategic planning such as electricity network developments. Plans for 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. But it has been historically weak. In the past, plans for extension of the network into new areas were based on fairly rudimentary calculations such as the predicted ratio of industrial to agrarian demand in a location. Plans for upgrade of the network within an existing area were based on basic aggregate statistics available at divisional level with no ability to consider specific local issues or the particular design of feeders / transformers / wire that would be most efficient.

The technical "mid-across" data now available is far richer – again reflecting big data's variety functionality – drawing on the analyses provided by FEMU and the growing body of historical trend and more immediate performance data. Indeed, official approval and funding for network changes can now only be sought from Stelcorp's Board on the basis of MDAS data analysis. So for selection purposes, there is now more objective data to stand alongside the relative subjectivities of politics and consumer complaint. And for planning purposes, there has been some strengthening of the role of technical data.

Modelling of upgrades has been enhanced – providing a more quantified and objective basis for load growth projection and for thus identifying what improvements to plan, where, and how – though the modelling 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.

Overall, then, there has been some displacement of human activity within the planning process – the envelope of automation is expanding – but the final decision and action components of information value chains remain human.

Value: Overview and Frameworks

R-APDRP and associated systems and the advent of big data within Stelcorp has cost somewhere in the region of US$50m but has clearly been associated with both organisational and wider social value. Inside Stelcorp, the promise of big data is reflected in the greater volume, velocity and variety of data being gathered. There are still issues with the accuracy and reliability of data capture, and there has been greater progress in the upstream than downstream parts of the information value chain: Stelcorp has changed more in regards to data capture and storage for example than it has in regards to decision making and action. Nevertheless, downstream impacts are visible for field / operational, tactical and strategic activity; activities which are now undertaken faster and better than before. Big data has touched all key aspects of Stelcorp's activities; improving billing, fault and loss detection and rectification, and planning of network upgrades and extensions. It has strengthened both culture and practices of performance audit and accountability.

In terms of conceptualisation of organisational value, only fairly simple models emerged as relevant. The information value chain was used to understand the extent of progress towards value formation but it is as much related to implementation as value. It looks feasible to follow the three-V characteristics of big data, seeking to understand the particular value contribution of each characteristic. And value contribution can also be understood by the time-honoured differentiation of decision types: operational, tactical, strategic.

Looking beyond the organisation to wider social value, arrival of big data has been used to guide improvements in the distribution network that have helped make parts of that network more efficient and more effective. Though the coupling is rather loose, there is also at least an association between big data and supply growth with a doubling in the number of consumers during the R-APDRP period and related growth in network capacity and assets. Stelcorp and Digicorp have both won plaudits at national level for their achievements, and been cited as exemplars of best practice.

But . . . stepping back to the broadest scope of value, the intention of R-APDRP was to reduce AT&C 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:

- AT&C losses – the aggregate measure of difference between units of power supplied and revenue collected – dropped significantly in 2010/11 but then rose back up to its earlier levels and has remained roughly static at one-third. 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 R-APDRP period without so far demonstrating stability of improved performance.

As discussed next, 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, Politics and Big Data

As explained in the literature review, the outcome domain investigates the broader implications of big data on development. This includes issues that emerged quite strongly from the case study: politics and power. This section reviews these, looking first at the intersection of big data and wider politics, and then at big data-enabled shifts in power.

Big Data and The Politics of Electrification

Why should it be that a big data initiative specifically introduced with the aim of cutting power losses has so far failed to achieve that goal? An explanation begins with disentangling the components of AT&C. First, there are the technical losses from resistance/heat and faults. Respondents, internal reports and reports of the Electricity Regulator all aver that the innovations described above are reducing such losses. Installation of online meters for bulk consumers is also said to be reducing losses, though the evidence base for this is weaker and sometimes contradictory. 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. This has been driven on by strong strategic emphasis from the central Ministry of Power on growth of rural electrification, and by a new State Government policy of 24 x 7 power for all which has impelled extension of the power network even into "shallow-demand" rural areas. 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. Compensatory growth in rural losses coincided with the electoral cycle in Janakari State. 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. This interconnection between politics, elections and free or low-cost access to electricity in rural areas is long-running and widespread throughout India (e.g. Dubash & Rajan 2001, Joseph 2010).

Stelcorp AT&C loss patterns follow a wider pattern 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 including "facilitating line losses" (Min & Golden 2014:624). As Min & Golden (*ibid.*) demonstrate, this appears to be effective: there is a positive and significant association between AT&C loss levels and likelihood of seat retention at election[4]. Previously, this would involve a mix of growth in supply to flat-rate, unmetered consumers and to unbilled consumers. With extension of metering to large numbers of consumers, the former path closes, and so non-payment of bills and illegal connections become the main routes.

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

[4] Specifically, "an increase in line loss of 10 points ... is associated with a 12 percent increase in the proportion of seats retained by the same party in a district" (Min & Golden 2014:624).

rationalised all this as a "social cost" that utilities had to pay, and which still contributed towards national development goals. But these costs are severe: they underpin huge financial losses for the power sector that restrict reinvestment and necessitate higher tariffs on paying consumers and higher state subsidies, and they underpin and reinforce political corruption. They may also underpin economic inequalities with historical experience, at least, suggesting it is often the more-affluent sections of rural society that benefit from this loss-led politicking (Dubash & Rajan 2001). Though we lacked hard evidence on this current project there were signs that this was repeated; that, as one might anticipate, 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.

Drawing on the earlier analysis, this could again offer itself to conceptualisation in terms of the design-reality gap model: a technology designed for some imagined private corporation of the global North that is objective and a-political in its decision processes, meeting the much messier and politicised reality of a public utility in the global South. Design assumptions about standardised processes, solely-economic objectives, particular ethical values, formalised decision structures, etc find themselves mismatching the lived reality of electrical and electoral politics in Janakari State.

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. In the comments of regulators and Central Government, one has a sense that the big data-driven technical and urban consumer loss reductions have allowed the State Government to "get away" with its political approach to rural electrification. The two effects of technical/urban loss reduction and political loss increase have roughly balanced one another out. The overall effect is that Stelcorp's performance has been disappointing but not so bad as to trigger some direct intervention by the regulators or by Central Government, and allowing for some credibility to claims in its strategic planning documents that losses in future years may be reduced.

Another conceptualisation that may help here is the epistemic view of big data. This draws on the idea that data creates a virtual model of real-world phenomena that exists in parallel to the physical version, but which may be more powerful in shaping decision making. For example, Taylor & Broeders (2015) refer to the "shadow map" created by big data: a virtual simulacrum of the real landscape that renders some elements more visible and some elements less visible than in the real-world equivalent. We can see this reflected in the way big data seems to have helped create a new imaginary for electricity in Janakari State, particularly within government. Interviewees explained that the pre-R-APDRP policy paradigm was one that saw electricity in terms of constraint: geographic constraint such that not all areas could be connected, and supply constraint such that "load-shedding" – regular blackouts and brownouts – was regarded as integral. Thus, for example, extension into new areas of demand would be selective: some areas would receive extensions but others would not; and extensions would be planned on the basis of pre-mandated load-shedding: a built-in plan that electricity would only be supplied for certain hours of the day.

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 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. Those design-reality gaps of introducing rational technology into a politicised context have thus led big data to reverse its intended outcome: increasing rather than decreasing losses at least in rural areas.

This epistemic shift in imaginaries has been to the benefit of rural consumers; particularly the millions who have been connected to the electricity network for the first time during the 2010s. But the big data landscape – the virtual equivalent of the real landscape of Janakari State – remains uneven. It offers greater presence to those entities, entirely in urban areas, which feed in real-time data; and it offers only a lesser presence to assets and consumers in rural areas for which data is time-lagged and also more error-prone. The extension of electricity to rural areas is a massive offset to any data-enabled inequalities but they nonetheless must be acknowledged: the slower, lesser, and poorer-quality decision making for rural areas that inevitably ensues from inequalities built into the current data infrastructure. For politicians, of course, this inequality may be beneficial. The rural areas in which the majority of the electorate lives may be rather weakly represented on the shadow map, but that makes it harder for Stelcorp – FEMU particularly – to exercise data-driven control over these areas.

Big Data-Associated Changes in Locus of Power

There are many ways in which changes in power associated with big data could be mapped. Since data is – and was discussed by some of our respondents explicitly as – an organisational resource, then one approach we identified was the resource dependency perspective. This argues that the dependency of an organisation on a resource can be assessed on the basis of three factors (Pfeffer & Salancik 1978) which we here describe as:

- Criticality of the resource: "the extent to which the organization requires it for continued operation and survival" (*ibid.*:45).
- Control of the resource by some internal or external entity, defined in terms of possession, access, use and regulation of the resource.
- Commutability of the resource source: "whether the focal organization has access to the resource from additional sources" (*ibid.*:50) and could therefore replace or substitute current resource provision.

We can firstly analyse this in relation to data systems. Previously, Stelcorp was responsible for its own data and its own data systems: it had sovereign control over them. Given the neo-liberal flavouring of power sector reforms in India, and the encouragement – verging on requirement – for public-private partnerships in relation to R-APDRP (TetraTech 2012), this is no longer the case:

- In relation to criticality, the move towards a data-intensive development model has made data more critical than previously, and Stelcorp could not continue functioning in any viable form without the data contained within the MDAS that Digicorp operates for them.
- In relation to control, *de jure* ownership of all R-APDRP software and data lies jointly between Central Government and Stelcorp. But *de facto* ownership lies in the hands of Digicorp; for example, through its operation of MDAS and the ERP application. Hence Stelcorp IT staff complain about their lack of control over Stelcorp data and software, their dependency on Digicorp, and their diminution to just a clerical role. There are also 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. Digicorp and Stelcorp share access and use and the setting of internal rules and regulations around the data system; but Digicorp retains control of a key related resource: knowledge of the design, construction, operation and maintenance of the data systems (so that, for example, Stelcorp engineers stating they are unhappy about the speed with which data system problems are fixed, can do little about it).
- In relation to commutability, there is a grey area. In theory, Digicorp could be replaced by another external provider or even possibly by insourcing. But in practice, it would be very difficult to do this because of the criticality and control dependencies. We thus see a significant asymmetry of power – as seen in other PPPs where public sector bodies have felt unable to hold private "partners" to contractual obligations such as penalty payments (e.g. Pollock et al 2001, Froud 2003) – such that the private partner is locked-in. Given continued billing for the services it provides, there is no incentive for Digicorp to surrender these dependencies and – as some staff request – to hand over control of the data systems to Stelcorp.

All this adds up to a substantial shift in data-related power from the public service provider to its private sector IT contractor.

Alongside this shift from public to private – which has also been a shift from inside to outside the organisation – we can also identify an upward shift in power from labour to management, as seen in the case of meter-readers. With the introduction of online meters for bulk consumers and urban components of the distribution network, they are no longer required to read those meters. As a result, during 2013-2016, 40% of meter-readers lost their jobs. For the remainder, the writing is clearly on the wall – online metering will spread throughout the rest of the electricity network, and their jobs will slowly but steadily be automated out of existence. For those that remain, the data they collect is less critical than previously, as it forms a declining proportion of all meter data; they have less control when reduced to just capturing data on their CMRIs (they barely own and access this data and do not use or regulate it); and they face actual replacement by online meters. As a result, Stelcorp managers are decreasingly resource-dependent on the meter-readers and power has shifted away from the latter and towards the former.

Reinforcing this upward shift in power has been the change within Stelcorp's lines of management accountability (see Figure 10). Previously, managers at divisional and zonal levels would be accountable to their immediate superiors, typically within that level: it was those superiors who collected performance data on the managers and negotiated its implications, and those superiors who held power over their junior managers. With the introduction of big data and FEMU, this has now changed.

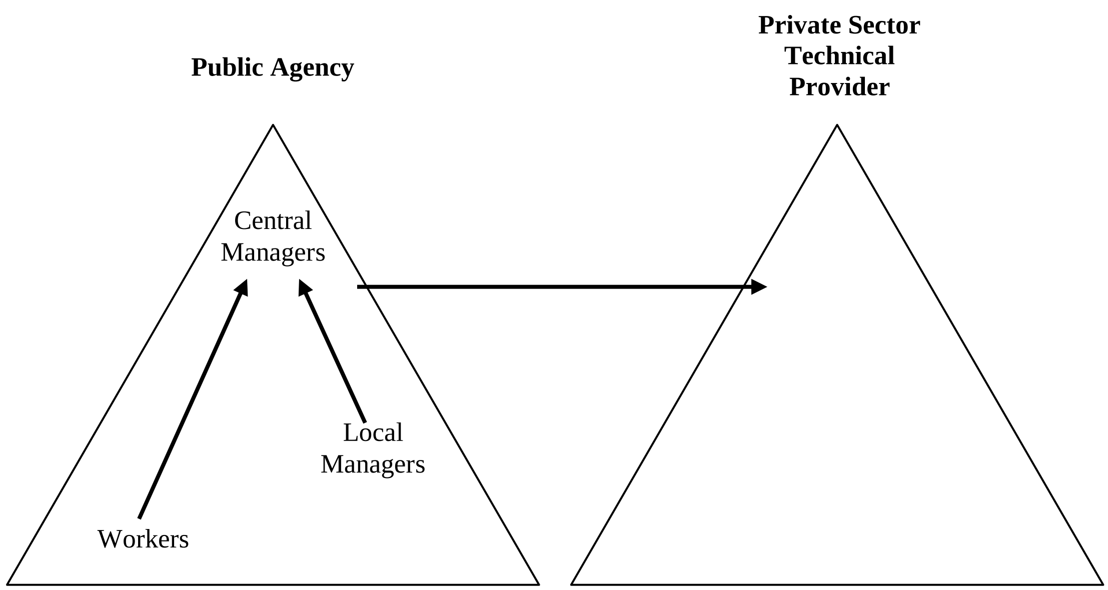


Figure 10: Power Shifts Associated with Big Data

Previously, data within Stelcorp was relatively “sticky”; tending to be restricted to localised enclaves within the organisation. Since the data-related innovations, this is no longer the case: even quite detailed data can be moved to virtually any part of the organisation. This opens the possibility for a number of structural and procedural innovations. For example, it would allow for peer-assessment of performance or the creation of performance-based communities of practice. However, the model chosen within Stelcorp is that of the Panopticon: Jeremy Bentham's 19th century ideal-type prison; a constructed technology that allowed a single central guard to observe the activity of all prison inmates (Tufekci 2014). FEMU is Stelcorp’s corporate guard.

We attempted to understand this in terms of the resource dependency approach, but this proved too limited to encompass the evidence and analysis we developed. Instead, we turned to a union model of different components of power (see Figure 11): resource, practice, institutional, positional, and epistemic power (Heeks forthcoming).

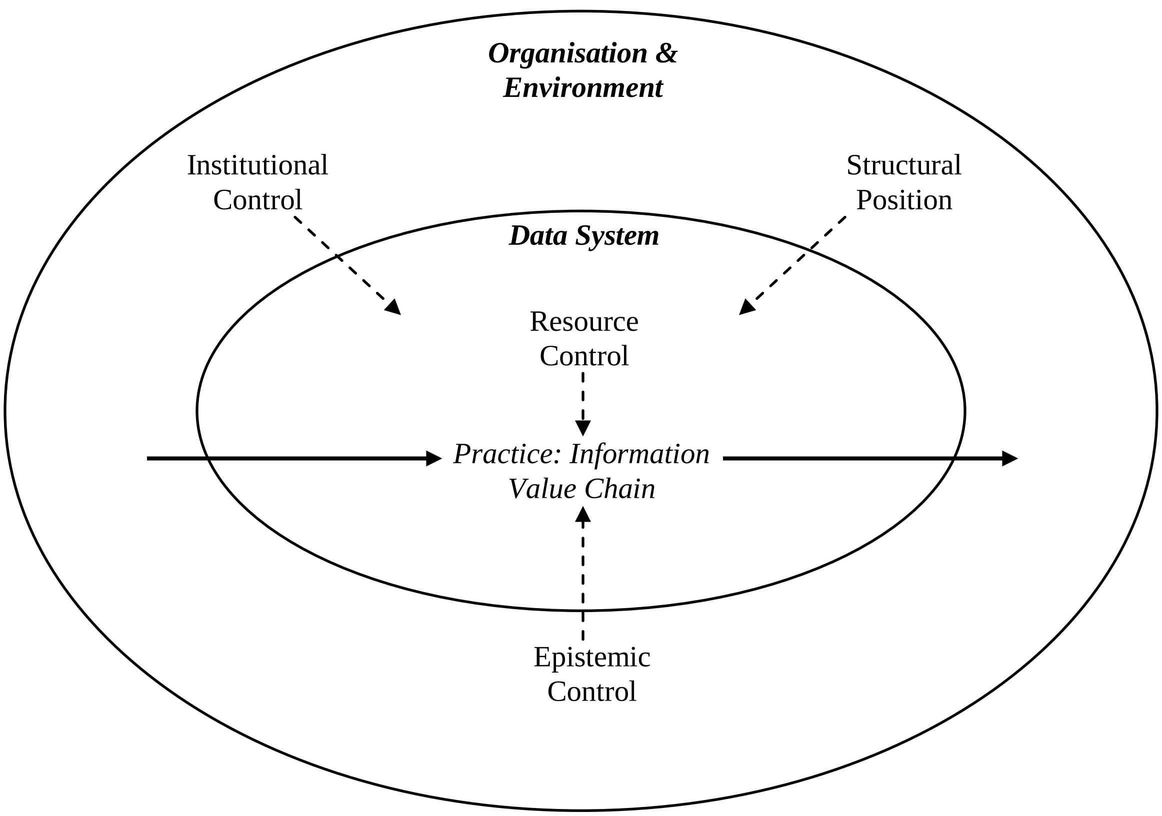


Figure 11: Sources of Data System-Related Power

FEMU can be seen to have strengthened all five main sources of power by gathering unto itself:

- the resource power of data – seeing all that goes on within Stelcorp (at least to the extent reflected by current online data) and being able to drill down through zonal and divisional data to individual assets;
- the power of new practices via the regular performance meetings it holds with Stelcorp managers;
- the institutional power of its authority such that managers now significantly see themselves as accountable to FEMU, and a looser institutional power in introducing more of an audit and performance management culture within Stelcorp.

In terms of positionality, it now has a direct structural relation to almost all Stelcorp managers. It has also inserted itself into the relationship between Stelcorp and external agencies, with its audit reports now serving as the main communicative mechanism to state and central audit and regulatory authorities. As always with such reports but bolstered by the presence of big data, they serve epistemically to present an objective and technocratic image of the organisation to the outside world (an image that belies at least some of Stelcorp's messier realities). FEMU's reports therefore increasingly form the basis for at least official internal and external discourse around Stelcorp's operations.

Finally, one can note a shift in locus of power that has not occurred: a shift from public corporation to citizens and citizen groups. Open data initiatives that provide datasets online have shown considerable promise in empowering external stakeholders to hold public organisations to account, to press for service improvements, and to correct data inaccuracies (Kitchin 2014). But there is 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.

DISCUSSION AND CONCLUSION

Emergent Issues

A number of emergent issues related to specific domains: those particular areas of big data and global development identified for and from the literature review. In relation to implementation, the case study echoed to some extent the narrative of the literature by exposing both technical and human challenges that were preventing Stelcorp's big data system from functioning to full capacity: unreliability of digital hardware and networks; problems of data standards; and data quality being undermined by human actors not working as intended, by clashes between working processes inscribed into different systems, and by problems with data coding. However, while these hamper the operation of big data – requiring ongoing human intervention, introducing some uncertainty into big data-based decisions – they do not undermine it. This is a big data project that has moved beyond proof-of-concept, beyond pilots and prototypes, to organisational-scale operation.

In relation to value, the main emergent issue was the "curate's egg" of impact that was good in parts but in other areas had not changed much. Billing was more accurate but not faster; fault/loss rectification was faster but not as fast as it could be; managers and the organisation were more internally- and externally-accountable but this may be more potential than fully-realised; tactical and strategic decision making were better and faster but do not make full use of big data's potential. Notwithstanding these limits, evidence on value emerged that was well beyond the more tentative or hypothetical evidence base that was typical of the literature to date.

In relation to outcome, a better evidence base also emerged to confirm issues proposed in the literature to date, with big data having impacts on power and politics. But this was not politics in isolation. Instead, an emergent theme was that big data in developing countries represents the intersection of the technical and the political; albeit this was framed in multiple ways. First, there was a techno-political context with the reforms to India's electricity sector intertwining new technologies and a new political agenda. The new technologies were all in some way data-related – linking to some part of the information value chain – but they were delivered by political innovation. There is no necessary relationship between neo-liberalism and big data, between a growing role for the private sector and big data. But that connection has been dominant within development (Taylor & Broeders 2015) and was dominant in this case: the principal – indeed only – narrative was one in which the private sector is required in order to deliver big data.

Given this techno-political coupling of neo-liberalism and big data at the foundation of innovations, there are some unsurprising shifts in the locus of power enabled by this new technology. The public sector organisation becomes more resource-dependent on its private sector partners as a result of big data; thus transferring power to those partners who own the data systems and data systems-related capabilities: those who control the information system value chain. Managers, as representatives of capital, become less resource-dependent on at least one section of labour – the meter-readers; thus transferring power from those workers. And there is a further concentration of power as big data enables culture and practices of accountability and performance management to be centralised. Those in central management who can access the data and have data-related capabilities of decision-making and action – those who control the information value chain – were therefore empowered.

But this big data case study is not solely a story of techno-political change: of a neo-liberal political agenda driving technical changes which help instantiate that political agenda. It is also a story of techno-political reproduction given that neo-liberal impulses in India sit alongside deep and long-standing clientelist politics. Thus while big data has been harnessed by the private sector it – or at least its imaginary and its potentiality – has also been harnessed by the incumbent political party in Janakari State. Big data has partly encouraged, partly enabled a massive expansion of electricity distribution; much of which delivers electricity for electoral benefit but not financial benefit.

Conceptual Frameworks

There were individual frameworks that emerged as helpful in categorising specific findings, such as the hierarchical model of different decision types, or potential models of human intermediaries. But four conceptualisations emerged as having a wider analytic value.

The three-Vs (or variants) notion of big data is so trite that it can be overlooked. But we saw here how – with the additional characteristic of granularity – it can be helpful in understanding the value of big data. Each of the data characteristics brings a different value, and it is relatively straightforward to tease out those values.

The information value chain found a number of uses. Our four-domain classification proved useful not only in categorising the literature on big data and development, but also in providing the overall structure for our case study analysis. The underlying value chain was useful in explaining the nature of different innovations taking place: and this could be extended by breaking the "data" component into constituent processes of data capture, input, processing, storage and output (Heeks 2006). The chain also helped in tracking implementation and value-creation; particularly via the notion that these had so far progressed further in the "upstream" data-related processes than in the "downstream" processes related to information, decisions and actions.

Like the information value chain, the design-reality gap model was primarily associated with analysing implementation but with some wider possible contribution. As already demonstrated (Gomez & Heeks 2016), its most obvious application is in understanding the nature and extent of barriers faced by big data projects. This provides an explanation for the extent of success or failure seen in such projects, and can be applied – as here – once a project is well under way or at earlier stages including pre-hoc assessment of readiness and risk. This enables application to post-implementation issues: helping explain why particular patterns of value and outcome were seen.

Turning finally to outcome, we found contributions from bounded models of power. The epistemic model – the idea of shadow maps and data imaginaries – helped analyse the power of knowledge systems and discourse. Specifically it traced the way in which perceptions and understandings of real-world systems changed as big-data-based virtual representations of these systems came to prominence. Resource-dependency theory helped analyse resource power; very specifically the power of data as a resource and the way this changed in moving from traditional to big data. But these were both sub-sets of broader models of power, such as that shown in Figure 11. These can be used to provide more comprehensive analysis of the way in which big data enables shifts in power, and to answer that most fundamental question of development impacts: "*Cui bono?*".

Conclusion

This paper has reported an exploratory and largely inductive case study of big data and development. It draws from one sector – the electricity sector – in one country; India. Being purely exploratory and based only on a limited base of evidence, one cannot conclude too much by way of generalisations. Instead, the main and realised intention was to identify a series of specific and generic issues worthy of further research, and a series of specific and generic conceptual frameworks that might be used in future to research those issues. From this foundation, the next stage can be more focused and deductive work that investigates the implementation, value and outcome when big data is introduced into a development context.

References

Arthapedia (2015) Restructured Accelerated Power Development and Reforms Programme(R-APDRP), *Arthapedia*, 23 Jul http://www.arthapedia.in/index.php?title=Restructured_Accelerated_Power_Development_and_Reforms_Programme(R-APDRP)

Bajaj, H.L. & Sharma, D. (2006) Power sector reforms in India. In: *PEDES'06. International Conference on Power Electronics, Drives and Energy Systems*, IEEE, New York, NY, 1-5

Benbasat, I, Goldstein, D.K. & Mead, M. (2002) The case study research strategy in studies of information systems, in: *Qualitative Research in Information Systems*, M.D. Myers & D. Avison (eds), Sage, London, 79-99

Blumenstock, J.E., Keleher, N.C. & Reisinger, J. (2016) The premise of local information, paper presented at *ICTD2016*, Ann Arbor, MI, 3-6 Jun

Dearden, A. & Rizvi, S.M.H. (2015) ICT4D and participatory design, in: *The International Encyclopedia of Digital Communication and Society*, R. Mansell & H. A. Peng (eds), John Wiley & Sons, Chichester, UK, 395-403

Dubash, N.K. & Rajan, S.C. (2001) Power politics: process of power sector reform in India. *Economic and Political Weekly*, 36(35), 3367-3390

Froud, J. (2003) The Private Finance Initiative: risk, uncertainty and the state. *Accounting, Organizations and Society*, 28(6), 567-589

Gomez, L.F. & Heeks, R. (2016) *Measuring the Barriers to Big Data for Development*, Development Informatics Working Paper 62, GDI, University of Manchester, UK

Hawari, A.A. & Heeks, R. (2010) Explaining ERP failure in a developing country: a Jordanian case study. *Journal of Enterprise Information Management*, 23(2), 135-160

Heeks, R. (2006) *Implementing and Managing eGovernment*, Sage, London

Heeks, R. (2015) Data-x development: what's in a name, *ICT4DBlog*, 16 Nov https://ict4dblog.wordpress.com/2015/11/16/data-x-development-whats-in-a-name/

Heeks, R. (2016) A research agenda for data-intensive development, *ICT4DBlog*, 18 Jul https://ict4dblog.wordpress.com/2016/07/18/a-research-agenda-for-data-intensive-development/

Heeks, R. (forthcoming) *Information and Communication Technology for Development*, Routledge, London

Heeks, R. & Renken, J. (2016) Data justice for development: what would it mean?, *Information Development*, prepublished, 8 Nov

Heeks, R. & Sampaio, P. (2016) *Big Data for International Development: Research Agenda*, Centre for Development Informatics, University of Manchester, UK

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: a review of promises and challenges, *Development Policy Review, 34(1)*, 135-174

IAEG (2014) *A World that Counts: Mobilising the Data Revolution for Sustainable Development*, Independent Expert Advisory Group, United Nations, New York, NY

IndiaMart (n.d.a) Secure Meter, *IndiaMart* http://www.indiamart.com/proddetail/secure-meter-4943619255.html

IndiaMart (n.d.b) Common Meter Reading Instrument CMRI, *IndiaMart* http://www.indiamart.com/pinnacleengineeringsolutions/common-meter-reading-instrument-cmri.html

Isafiade, O. & Bagula, A (2013) Efficient frequent pattern knowledge for crime situation recognition in developing countries, paper presented at *4th Annual Symposium on Computing for Development*, Cape Town, South Africa, 6-7 Dec

Jacobson, A., Milman, A.D. & Kammen, D.M. (2005) Letting the (energy) Gini out of the bottle: Lorenz curves of cumulative electricity consumption and Gini coefficients as metrics of energy distribution and equity. *Energy Policy*, 33(14), 1825-1832

Jahani, E., Sundsoy, P.R., Bjelland, J., Iqbal, A., Pentland, A. & de Montjoye, Y.-A. (2015) Predicting gender from mobile phone metadata, in: *NetMob 2015*, E. Moro, Y.-A. de Montjoye, V. Blondel & A. Pentland (eds), MIT Media Lab, Cambridge, MA, 110-113

Joseph, K.L. (2010) The politics of power: electricity reform in India. *Energy Policy*, 38(1), 503-511

Kitchin, R. (2014) *The Data Revolution: Big data, open data, data infrastructures and their consequences*, Sage, London

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

Mam, M., Joseph, A., Singh, A., Alam, M.P. & Popli, S. (2014) A case study on issues and challenges faced during strengthening of a distribution unit by reducing aggregate technical and commercial loss. *International Journal of Electrical and Electronics Research*, 2(1), 36-40

McDonald, S.M. (2016) *Ebola: A Big Data Disaster*, Paper 2016.01, The Centre for Internet and Society, Bengaluru, India

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

MoP (n.d.) *R-APDRP*, Ministry of Power, New Delhi, India http://www.apdrp.gov.in/Forms/Know_More.aspx

MoP (2015a) *Review Planning and Monitoring Meeting*, Ministry of Power, New Delhi, India http://www.apdrp.gov.in/Whats_New_Files/IPDS_RPM_12102015.pdf

MoP (2015b) *IPDS in Smart Cities*, Ministry of Power, New Delhi, India http://smartcities.gov.in/writereaddata/Smart_City_Presentation_IPDS.pdf

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

Pfeffer, J. & Salancik, G.R. (1978) *The External Control of Organizations*, Harper & Row, New York, NY

PLJ (2014) *Nowcasting Food Prices in Indonesia using Social Media Signals*, UN Global Pulse Project Series no. 1, Pulse Lab, Jakarta

Pollock, A., Shaoul, J., Rowland, D. & Player, S. (2001) *Public Services and the Private Sector*, Catalyst, London

Pradesh18 (2015) Fearing massive power bills, residents wary of late meter reading, *Pradesh18*, 22 Jan http://english.pradesh18.com/news/bihar/fearing-massive-power-bills-residents-wary-of-late-meter-reading-675713.html

Selvam, M.M., Gnanadass, R. & Padhy, N.P. (2016) Initiatives and technical challenges in smart distribution grid. *Renewable and Sustainable Energy Reviews*, 58, 911-917

Srivastava, L. & Rehman, I.H. (2006) Energy for sustainable development in India: linkages and strategic direction. *Energy Policy*, 34(5), 643-654

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

TetraTech (2012) *Electrical Power Distribution: Case Studies from Distribution Reform, Upgrades and Management (DRUM) Program*, Tata McGraw-Hill, New Delhi, India

Tufekci, Z. (2014) Engineering the public: big data, surveillance and computational politics. *First Monday*, 19(7)

Winkler, H., Simões, A.F., La Rovere, E.L., Alam, M., Rahman, A. & Mwakasonda, S. (2011) Access and affordability of electricity in developing countries. *World Development*, 39(6), 1037-1050

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

Yin, R.C. (2012) *Application of Case Study Research*, 3rd edn, Sage Publications, Thousand Oaks, CA

Appendix: Timeline of Data-Related Initiatives in Stelcorp

Start Date	***Type of Initiative***	***Agency Involved***
1997	Digitisation of consumer energy and revenue data	Stelcorp
2000	Installation of digital meters at feeders	Stelcorp
2004	Replacement of analogue with digital meters for end consumers	Stelcorp
2004	Sharing of aggregate digital data from divisions to zones	Stelcorp
2009	Installation of online digital meters at urban feeders and transformers; offline digital meters at rural feeders and transformers	Stelcorp, R-APDRP
2009	Creation of central data centre with offsite disaster recovery backup	Stelcorp, R-APDRP
2010	Creation of Internet backbone between central data centre and zonal / divisional centres	Stelcorp, Digicorp, other private partners
2012	Operationalisation of MDAS	Stelcorp, Digicorp, other private partners
2012	Operationalisation of SAP-ISU application	Stelcorp, Digicorp
2014	ID coding of customer premises via respective transformers	Stelcorp