

Report on Open Government Data in India

Version 2 Draft: August 19, 2011

Glover Wright, Pranesh Prakash, Sunil Abraham, Nishant Shah and Nisha Thompson
Centre for Internet and Society, Bangalore

Table of Contents

Executive Summary	1
Benefits of Open Government Data	1
Challenges	1
Envisioning Open Government Data for India	1
Introduction	3
Current Data Practices and the Status of E-Governance	3
Access to Networked Technologies	4
ICT Revolution	5
Political and Administrative Environment	7
Civil Society	8
Media	9
Open Data: Why Now?	10
Methodology	11
Policy Environment.....	12
Right to Information in India	12
Shortcomings of Reactive Disclosures	14
Copyright Policy	15
Standards and Interoperability	17
National e-Governance Plan	18
National Knowledge Commission Recommendations on e-Governance	20
Public Information Infrastructure.....	21
Case Studies	23
Government Case Studies	23
Civil Society Case Studies.....	26
Public-Private Partnership Case Studies.....	29
Civic Hacking Case Studies.....	32
Forms of Access to Data	34
Challenges	35
Data Unreliability	36
Semantic and System Interoperability	37
Cost of Data.....	38
Scalability	39
Indian Names	39
Indic Languages and Technology.....	40
Technological Efficiency versus Access.....	41
Privacy.....	42
Elite Capture of Transparency.....	44
Observations and Recommendations	46
General Observations.....	46
Recommendations	47
Annexure I: Survey of Organizations and Projects	50
Annexure II: List of People Interviewed and Interacted With.....	55

Executive Summary

This report looks at some of the landscape relevant to open government data (OGD) in India, starting from the current environment in government, the state of civil society, the media, the policies that affect it such as the Right to Information Act, the standards-related policies, e-governance policies, and the copyright policy. This report also looks at a few case studies from government, civil society organizations, a public-private partnership and profiles some civic hackers. It then examines some of the varied challenges to the uptake of OGD in India, from infrastructural problems of e-governance to issues such as privacy and power imbalances being worsened by transparency. Finally, it lays out our observations and some recommendations. It concludes by noting that OGD in India must be looked at differently from what it has so far been understood in countries like the UK and the US, and providing some constructive thoughts on how we should think about OGD in India.

Benefits of Open Government Data

Are there benefits to be found from OGD in India? Yes. First, it will benefit government itself, by reducing the burden of locating information—both for internal use, as well as for responding to RTIs—and streamlining its own information gathering and processing procedures. Citizens and the variety of civil society organizations in India currently working with government data—everything from electoral candidate data and legislation to municipal body phone numbers and public transport route information—will be able to create better tools and analysis as well as elevate the discussion with government. It will expose incorrect and outdated data, which the government itself is often not in an easy position to detect. Most importantly OGD can be seen as a step in delivering the promises of the Right to Information Act, and a step towards greater transparency and importantly, accountability.

Challenges

There are many challenges that must be addressed while opening up government data. Data collection and processing methods are not set up for quick release. Data itself is not standardized or in useful formats. Also services and applications are not looking at making more data open. There is also still a great need to build infrastructure so that information can be collected and distributed electronically.

Currently the entire infrastructure of information gathering, processing and sharing requires a great deal of improvement. There is insufficient standardization, and e-governance has, to a large degree, not translated to more open data or system interoperability. There is an ambitious project by the Public Information Infrastructure seeking to tackle this situation, though. Even if system interoperability is brought about by use of common formats and software standards, there is still the issue of semantic interoperability—e.g., different departments gathering different information under the same heading, or the same information under different headings—that can't be tackled as easily.

Since there is no written robust law on privacy in India, and data anonymization is seldom practised. Even with anonymization, privacy is still an issue because of nature of India's community focused culture (e.g., showing in which villages HIV-positive people are concentrated, even if individuals are not mentioned which would bring shame to the whole community).

The capacity of civil society organizations to make use of information if it is put up in machine-readable (and non-human-readable) formats will be limited. The linkages between technologists and civil society organisations need to dramatically increase.

There is also the issue of linking e-governance initiatives with open data principles. Currently government is more focused on creating usable applications with no policy on opening the data that it uses or creates. Given the move to more e-governance schemes it is important to stress that the data created by these applications have other uses other than what the government designates.

Decentralisation would empower urban areas and panchayats to look at data as an important governing tool. For instance a city collects property taxes which make up a small percentage of their total budget however, a city can't control the equation used to determine how much in taxes property owners. This ties their hands and provides no incentive to collect information on whether the current tax rate is fair or accurate. This is just one example of how decentralisation would allow cities and villages to better gather information to govern more effectively.

Lastly, it is essential to understand that most of the Indian experience is informal. Because there is a huge informal sector and relationships are valued more than data there are a lot of gaps that exist in type of information gathered. Even if all current data is available one cannot get a full picture of what is happening.

Envisioning Open Government Data for India

It is our belief that open government data in India cannot be as much an issue of providing data for mashing and allowing for innovative private-sector information products. Instead, it must be more about addressing the shortcomings of the Right to Information Act, and extending and fulfilling its promises of the transparency as partly envisioned by the Knowledge Commission, and perhaps moving towards accountability. Some reasons for this:

- The RTI movement has proved itself to be credible, well-organized, and in a position to effect change
- The Right to Information Act itself requires a large amount of proactive disclosure
- The Knowledge Commission's report on e-Governance itself talks about making available more governmental information and data to the public.
- Technologists in the social and political sector ("civil hackers") are few, and the existing civil society groups are in a better position to take advantage of any governmental data that is opened up. Just as such organizations should not be beholden to the government to provide data as a privilege, they must not be beholden to technologists to provide them access to the data put out by the government.
- Civil hackers need to be located, engaged with, and encouraged to work with governmental data. Existing communities around free and open source software and around open content (such as Wikipedians in India) are groups that could be engaged with with this aim.

Thus, while data mashing and private-sector information products must be allowed, they must more importantly be facilitated, encouraged, and in some instances, be performed by the government itself. The duty of the government cannot end merely at providing information, but must

extend to making that available in such as form that facilitates analysis and enhances offline usability.

The government is already looking at many of issues of information infrastructure, including an open data policy. Hence there is a need to help the government in this regard, including by gathering examples of data usage in India, and studying the best practices and the problems in implementations of OGD in other countries. To ensure the relevance of open government data, mechanisms have to be put in place to take its benefits to the common person and to marginalised communities, both by the government as well as by civil society organisations; putting up raw data will not suffice.

Introduction

In India, as in other countries, the relationship between citizens and government is increasingly mediated by information systems, and e-governance is clearly seen as the way forward for efficient delivery of public services. This tendency has only grown since the 1980s. Now, India is preparing to introduce a national Unique Identity (UID) project, (now rechristened Aadhaar, meaning 'foundation') which will provide every Indian with a unique identification number and a corresponding entry in a national biometric database,¹ as a cornerstone of e-governance initiatives. Therefore, it is not only necessary to examine India's ecology of government informatics, but also a very opportune time to do so.

The Indian national government has made clear that in coming years it intends to make publicly available much of its data. In its tenth Five Year Plan (2002-2007),² the Indian government announced its intention to become a 'S.M.A.R.T.' (Simple, Moral, Accountable, Responsible and Transparent) state. Which seeks to use technology to improve systems and stem corruption, however there is very little in the Eleventh Ten Year Plan that requires an open database created to keep programs accountable. Several types of data bases are mentioned but are for internal use or are heavily protected because they include sensitive material. This has led to many e-governance initiatives,³ but few of them have resulted in publicly-accessible databases. And fewer still of those publicly-accessible databases are 'open' in terms of data re-usability (technologically, in terms of machine-readability and openness of formats), data re-usability (legally), easily accessible (via search engines, for persons with disabilities, etc.), understandable (marked up with annotations and metadata).

Right now the cornerstone legislation to move information into becoming open data is the Right to Information Act (RTI) of 2005. The Act initiated one of most important changes in the citizen-government relationship in India since its Independence. This legislation is applicable throughout all of India, and is a landmark in setting out a clear political agenda of transparency, signalling a shift from the opacity promoted by the Official Secrets Act. The RTI Act's section 4 actually requires agencies to proactively disclose information and data and can be a good base for open data schemes in the future.

This report situates the current move towards open government data in India in the context of the country's growingly sophisticated information and communications technology (ICT) practices as well as the Right to Information Act. It relies primarily on conversations—both on the record and off—with government officials, businesses, civil society organizations, and individual activists. For background it relies on a review of the literature relevant to OGD and RTI generally, to present a snapshot of where India stands now in respect to OGD, and to predict where it is likely to go in the near future. It seeks to understand what "open government data" means in an Indian context, and what effects institutionalized open data practices and ideas might have on Indian society. Finally, it suggests certain technical and policy strategies for developing, promoting, and implementing, and maintaining a robust open government data policy in India.

¹ 'Unique Identification Authority of India', available at <http://uidai.gov.in/> (Last visited 29 September, 2010)

² 'Tenth Five Year Plan of India', Planning Commission of India, available at <http://www.planningcommission.gov.in/plans/planrel/fiveyr/welcome.html> (Last visited 29 September, 2010)

³Some of the more important ones amongst these are detailed in the section below on the National e-Governance Plan.

Global Open Data Perspective

What needs to be kept in mind is that the global open data movement is promoting a specific vision of governance. Open data creates a way for people to interact with government by elevating communication techniques, improving services with better applications and using data to improve decision making, policies and procedures. This is based around the idea that governance is a group activity that should involve the public and in order for the public to participate they need to be properly informed not only with information but also with data that they can interpret freely instead of having only the government's side available. In order for this to happen it is essential that there needs to be a prevalence of information technology and adaptation. How this is being achieved primarily in the west (i.e. US and UK) is by advocating for open data standards and methods. Data itself is a way to this end and that is where government should start.

In order to accomplish this the Open Government Data Principles⁴ were created by the Open Government Working Group in 2007 to set a standard of what open data looks like and to establish what useful data formats are.

"Open Government Data Principles

Government data shall be considered open if it is made public in a way that complies with the principles below:

1. Complete

All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations.

2. Primary

Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.

3. Timely

Data is made available as quickly as necessary to preserve the value of the data.

4. Accessible

Data is available to the widest range of users for the widest range of purposes.

5. Machine processable

Data is reasonably structured to allow automated processing.

6. Non-discriminatory

Data is available to anyone, with no requirement of registration.

7. Non-proprietary

Data is available in a format over which no entity has exclusive control.

8. License-free

Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.

Compliance must be reviewable.

Definitions

1. "public" means:

The Open Government Data principles do not address what data should be public and open. Privacy, security, and other concerns may legally (and rightly) prevent data sets from being shared with the public. Rather, these principles specify the conditions public data should meet to be considered "open."

⁴ 'Open Government Data Principles', available at https://resource.org/8_principles.html (Last visited 27 June, 2011)

2. "data" means:

Electronically stored information or recordings. Examples include documents, databases of contracts, transcripts of hearings, and audio/visual recordings of events.

While non-electronic information resources, such as physical artifacts, are not subject to the Open Government Data principles, it is always encouraged that such resources be made available electronically to the extent feasible.

3. "reviewable" means:

A contact person must be designated to respond to people trying to use the data.

A contact person must be designated to respond to complaints about violations of the principles.

An administrative or judicial court must have the jurisdiction to review whether the agency has applied these principles appropriately."

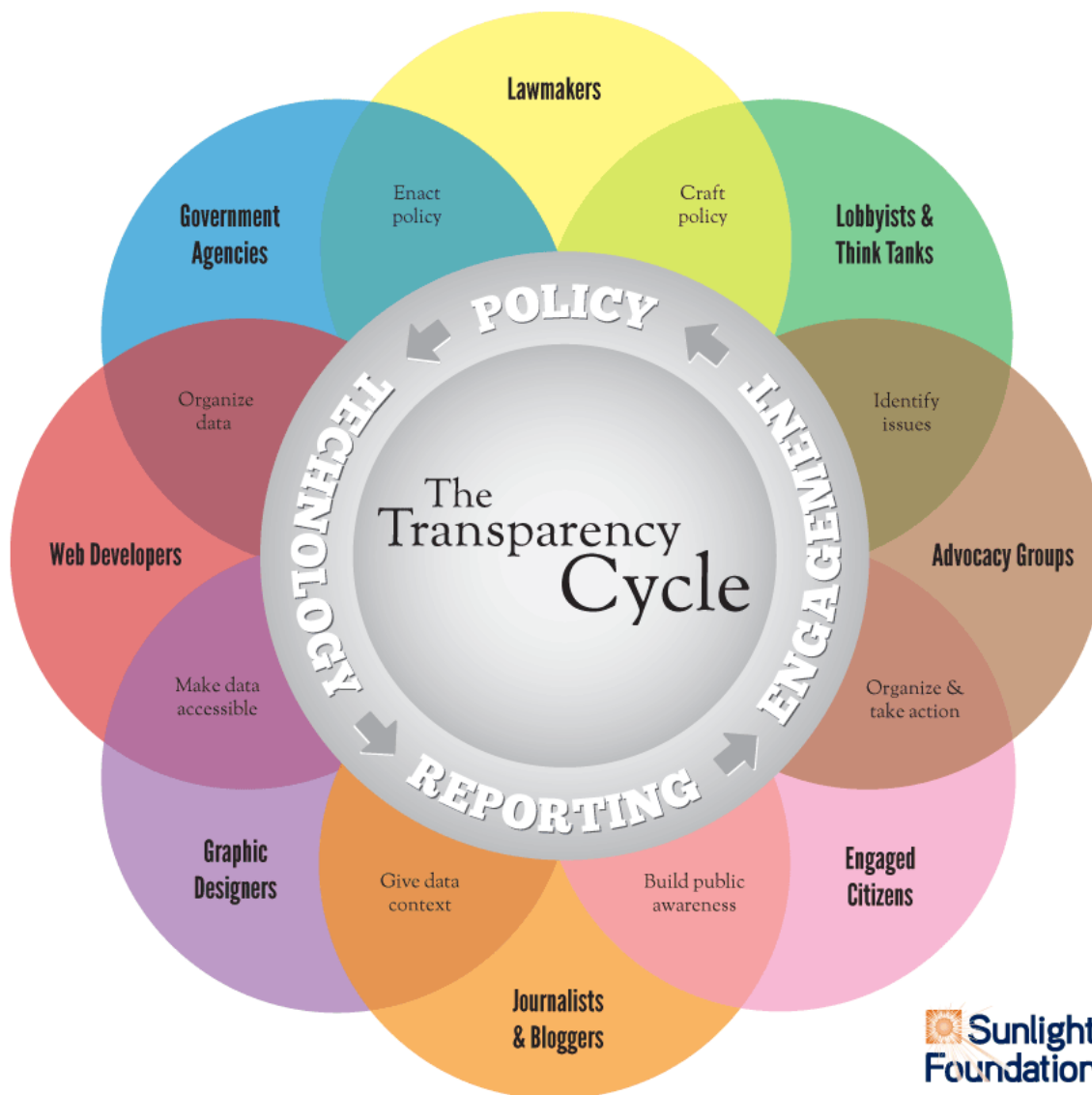
Sunlight Foundation (a US based NGO that is dedicated to transparency in government) later added a two more principles. The 9th around permanence: that data has to be available for a significant period of time and not just put up then immediately taken down. The 10th around usage costs: that charging fees for data prohibits people from accessing it and that this data has to be collected anyway so fees cause no significant difference.⁵

These principles tend to be echoes in open data policy in several countries such as Georgia (<http://transparency.ge/en/ten-open-data-guidelines>) and the UK (<http://www.information-age.com/channels/information-management/news/1263943/government-proposes-open-data-principles.shtml>) While the exact terms might change from place to place there is a significant amount of overlap.

The 8 Principles are the usually starting point for open data in developed countries like the US and the UK and other similar nations. The goal after the principles is better shown in the following cycle of transparency created by the Sunlight Foundation.⁶

⁵ '10 Principles of Open Data' Sunlight Foundation (<http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>)

⁶ 'Cycle of Transparency' Sunlight Foundation (<http://sunlightfoundation.com/TransparencyCycle/>)



While many considered this a simplification of the system it is clear that there are lot of players with over lapping responsibilities and functions who can work together to achieve transparency and openness. While this is a theory it is important to understand it so that the differences with the Indian experience can be clearly seen.

In India this cycle will not work mainly because there are not enough private players to interpret and make that data useable, access to technology has to become more prevalent and feedback mechanisms with government are limited. The government will have to be responsible for a great deal of these functions plus building a robust ICT infrastructure and better collection methods.

Current Data Practices

The Indian government collects a wide variety of data at the national, state, and district levels. The chief problem with data collection, according to high level government officials in New Delhi, is that India still lacks automation of processes at all levels of government. The more local and specific the data the less robust is the collection method. So while the data is being collected and ultimately made available in—or at least translated into—digital form, it is a time- and resource-consuming project, particularly below the national level where technology practices are either not in place or have not fully been cemented and the collection itself is of the first degree.

In general, data is collected in a systematic and timely fashion; the problem is the lack of consistency in the various terminologies and methodologies employed by different authorities. Moreover, data older than about 10 years, in particular, is likely to be on paper—though most public authorities at the national and state levels seem to be in the process of digitizing their collections if they have not already.

It is unclear in what formats most government data is stored. In general, where government officials have addressed this question, they have indicated that data is collected and stored in “spreadsheets,” sometimes identified as Excel files. However, there is a lack of understanding of what constitutes open versus proprietary standards and why one would choose (or would be forced to choose) one over the other. Also, importantly, when data is made publicly available, even though the government maintains it in machine-readable formats, it sometimes provides only scanned PDFs or otherwise tough-to-manipulate forms.

Data is housed in silos with very little desire to share it. The lack of standards is especially noticeable at the city and panchayats levels where information is not collected in systematic ways and there are huge issues of accuracy. Data is seen as an internal tool that can't fall into the wrong hands. What is collected is also limited and unknown. If they do have it is rarely in one place and has to be collected from several offices. This means organizing data itself will require a great deal of internal restructuring at the local levels. There are a few CSOs that have taken up the cause to collect data.

The current political environment relative to open government data is probably most driven by the still-influential right to information (RTI) movement and the good press which the RTI Act has garnered both in India and abroad. High level government officials believe that the public is hungry for more transparency and accountability, and that the RTI Act was a harbinger of an even more open government to come. Midlevel government officials interviewed at the national and state levels, however, seem to think that the RTI Act has done the job of making government transparent and thus that government is (mostly) already as open as it needs to be. A common refrain heard is that “It's all already available.” These officials admit that there is always room for improvement but offer no vision as to what sort of improvements might be necessary or when they should be implemented. Essentially, it seems that those at the highest levels of government feel empowered to push for a more robust information regime, including open government data, but are wary of pushback from the bureaucracy, which they and civil society say is threatened by the power/knowledge with which open government data endows citizens. But most everyone interviewed said that the movement in India is towards greater and more accessible disclosure of government data.

Political and Administrative Environment

India may actually be at a particular stage in its political cycle that would make opening government data an advantageous move. Especially in light of the success of the RTI movement and the continued recognition/popularity of the ensuing act, it is hard to imagine that any politician would choose to advocate for less openness.

The winners would likely be the politicians and, if a majority of proponents of RTI in India are

to be believed, the villages and districts which OGD would eventually empower to make their own decisions as opposed merely to following policies set out by national and state governments. Those supporting open data in India tend to see it in almost a Gandhian sense, as enabling villages to regain the power of self-determination that they have lost in the emergence of an industrial, urbanized India. The losers, certainly, however one looks at it, are the bureaucracy, who must cede some power both to the top that mandates data be made available and the bottom that uses that data both to hold them accountable and to make decisions without relying so much on the bureaucracy to guide them. The idea is that open data in India, if properly implemented—that is, if it is accessible and can be made understandable to “the common man”—will be a hugely democratizing force. The problems, of course, are numerous—access barriers, technology, reading literacy, language challenges, etc.—and no one believes that open data can be seen as an end unto itself, but rather as one vital part of an emerging knowledge ecology the other pieces of which must also come together for any real progress to be made. In explicitly political terms, though, they seem to view it as the most powerful of these parts, impacting society as a whole as RTI did the individual.

A senior source in a governmental department, who opted to speak anonymously, suggested that opening significant amounts of government data, while politically problematic at some levels of the bureaucracy, will be less of an issue once the country has succeeded in automating government processes as part of its national e-governance plan and mandating that bureaucrats use electronic systems instead of paper. Moreover, the source and others note that once the country has automated its processes, opening the data is of negligible expense. It also bears noting that in India the current proposal for an open government data scheme is tied up with the proposal for a national identification number, similar to a Social Security number, for which data warehouses and the necessary infrastructure to run a nationwide distributed database system have already been planned and in conjunction with would run India's proposed “Public Information Infrastructure.”

Openness does not seem to be a high stakes issue within the government; in general, almost everyone is at least verbally in favor of providing more information to the public. Those interviewed, however, almost unanimously asserted that the bureaucracy will provide a roadblock to open government data, both because they will feel as if they are losing authority and because it will make more work for them in the sense that they will have to dedicate greater time and effort to collecting data and ensuring its accuracy. They also noted, importantly, that India's bureaucracy has become quite adept at inhibiting policies that it doesn't like—that is, Indian bureaucrats know well the art of delay, even when a policy is mandated from the top. But people feel that as younger civil servants move into government—often with a more pronounced nationalism than the previous generation's that translates into a progressive vision for India—such opposition will decline. As for data that was not collected with the intention of being released, there will doubtless be opposition for various reasons; one state-level agriculture official, for example, remarked that he could not release detailed data on plant parasites because it might cause panic among his state's farmers and unduly affect global markets. Most powerful, though—and open data advocates believe this—is probably the fact that the more information is made available, the more power people at the lower political levels will have to take control over their communities and the less deference they will have to show for the entrenched bureaucracy. (For all these reasons, a mandate from the top will probably be necessary for an open government scheme that aims to release data

not originally meant for the public.)

Different accounts have been given by different officials as to the proficiency of the public administration with technology. While Shailesh Gandhi, an Information Commissioner at the Central Information Commission believes that the majority of government employees just can't work with computers. Others believe that the issue is not so much about competence but about willingness to engage with technology—many officials, for example, though they are provided with and understand e-mail, will only respond to telephone solicitations. The government has engaged with technology at all levels, for example providing information on services via text message and hosting information portals on their websites. As big a problem is society's inability to make use of the technologies employed by government, especially outside of urban areas. Beyond technology is the idea that much of the Indian experience is informal. Data collection points currently don't look at how much time bureaucrats and elected officials spend on informal tasks. Information on all aspects of what people do has to be looked at. While technology use has to be adopted it also has to be accepted that informal parts of their job that focus on relationships should become formal using technology is improve efficiency.

Finally, no one in India doubts the capacity of the Indian government to do what it sets out to do. The problem, said one high level official, lies in convincing it to act in the first place. E-government seems to have penetrated all levels of government, though, in awareness if not in actuality, and there seems to be little opposition to the idea of e-government generally. It is likely that a favourable political environment such as the one existing now may provide the nudge necessary for the bureaucracy to fully embrace the technologies that most sources within the bureaucracy admit make their lives easier, although at the expense of some power.

Access to Networked Technologies

A report by Boston Consulting Group found that while only 7 per cent of India—about 80 million people in a total population of about 1.2 billion⁷—has Internet access, the Internet is expected to reach 19 per cent of the country's population by 2015. Partly this will be a result of the government's plan to roll out a nationwide optical fibre network beginning this year, providing Internet access to all regions of the country.

About 60 per cent of India has mobile phones, but there are only about 2 million active users of mobile Internet services—less than 0.1 per cent of India's population. This fact seems to lie in stark contrast to the assertions by many both in government and the private sector that the future of the Internet in India is on the mobile. Although there are many anecdotes, for example, of farmers receiving agricultural and environmental data on their mobiles via text message, it is unclear to what extent these discrete stories are genuinely illustrative of a broader phenomenon. A recent study does indicate, however, that about 13 per cent of rural users currently seek farming information online, including information on fertilizers and pesticides.⁸ Unfortunately, the same study—a joint undertaking of the Internet and Mobile Association of India (IAMAI) and IMRB International—indicates that about 84 per cent of India's rural population either lack familiarity with

⁷http://www.google.com/publicdata?ds=wb-wdi&met=sp_pop_totl&idim=country:IND&dl=en&hl=en&q=population+india

⁸<http://www.ibtimes.com/articles/60385/20100908/interent-users-email-brici-countries-rural-india-online-application-railway-online-ticket-booking-re.htm>

the Internet or are unable to take advantage of Internet technologies.⁹ More positively, the study finds that rural Internet users are set to increase by 30 per cent from 3.3 million in 2008 to 5.4 million in 2010. It also indicates that the great majority of rural Internet users—more than 70 per cent—currently access the Internet through either the country's Common Service Centres or cyber cafes, good news for a public knowledge program seeking to deliver information through such centres.¹⁰

People have to become comfortable with using technology as a way to get and share information then they can easily adapt to using government and private applications. This will shine a light on data needed to create better services which will also create more demand for information from government.

ICT Revolution

As one of the key transformative factors in a globalizing India, the advances in ICTs have transformed everyday life and how people interact and interconnect with each other, communities, states, and markets.

The so-called "ICT revolution", as noted and debated by scholars, politicians, and policy-makers, has had an inordinate effect on economies and societies, leading to what has been termed a "global shift" (Cerny 1995). Erwin Alampay, in his introduction to *Living The Information Society in Asia* (2009) looks at Cerny's ideas of the 'global shift' as marking two fundamental alterations to the global political economy. He says,

First, there is the movement from an industrially-based international economy to one that is information- and knowledge-based. For some, these changes signal the emergence of the 'Third Industrial Revolution' which is both transnational in character and based on post-Fordist regimes of accumulation. Second, the ICT revolution is said to have profound positive and negative social, political and economic consequences that can become factors in determining development and underdevelopment. As such, ICT and its management have become a new rhetoric of development (10)

Richard Ling (2009), has written about questions of technology and access over a period of time. In his latest essay, he emphasises that 'interaction between technology and society'(14) has been at the core of most debates around emergence of new technologies. In his essay "What would Durkheim have thought?" Ling compares the kind of changes that were brought about with the Industrial Revolution and the changes ushered in by the Information Revolution. He sides with the position the James Beigner (1986) took in his book *The Control Revolution* and argues that 'we have not really experienced an information revolution. Rather, the increasing demand for control of ever more complex systems has resulted in a parallel, but perhaps somewhat lagged development of information systems' (Ling, 15) In exploring the various approaches to Technology-Society Interaction, Ling finally posits three crucial questions that he sees as necessary to be answered by ICT4D practitioners:

First, what characterizes the adoption process at the personal level? Second, after adoption has

⁹<http://topnews.com.sg/content/24865-84-per-cent-population-rural-india-unacquainted-Internet-study>

¹⁰<http://www.ibtimes.com/articles/60385/20100908/interent-users-email-brici-countries-rural-india-online-application-railway-online-ticket-booking-re.htm>

taken place, how does the object or service become integrated in our daily lives? And finally, how is the object or service interpreted by others after it has been adopted? (17).

Ling's work is useful to reflect on the technology-infrastructure practices which have been at the centre of much policy and governance debates around the world. He stands out as a strident voice that refuses to look at technology as the panacea to all problems and instead insists on radically re-investigating the relationship that technology has with the various interlocutors.

In the same anthology, Jean-François Doulet and Shang Dan look at how urban dwellers in China integrate mobility in their everyday life with the help of ICTs. In particular, they look at how access to information provides people with new spatial strategies, and allows them to become more mobile and confident in exploring the real world. Drawing from the work of John Urry and Elizabeth Buchannan, they posit, 'that since new mobilities produce and develop extensive and far-flung social connection, it is necessary to examine topologies of such social networks and especially the patterning of weak ties that generate 'small worlds' amongst those apparently unconnected' (52) In addition, they look at the changing socialization patterns among new urbanites, from small social circles based on deep personal relationships, to larger social circles based on common interests. They end their study by posing two questions which they see as crucial to understanding technology interaction and how it affects social conditions (both spatial and relational): 'how (do) people assess their own mobile living arrangements? Is becoming mobile mainly an autonomous decision or a forced situation? (55).

Even within contexts where ICTs are not the primary object of discourse, the understanding of access remains the same. In a report on Moving Toward Knowledge Based Economies (2007) by the Asian Development Bank, ICTs are considered as one of the four pillars of growth and development in Asia. The report recognises the various forms of e-governance sites which are right now being run in India and looks at them as nodes of open data production and dissemination: "[R]ural Internet kiosks, community e-centers, e-healthcare, geographic information systems (GISs), dairy sector applications, teacher training, online agricultural systems, wireless local loop solutions, databases of rural innovations, and other services targeted at women and children. In the realm of public service, e-Government projects include online delivery of land titles, land and property registration, and empowering dairy farmers through a dairy information and services kiosk". (33)

The report suggests that the emergence of ICTs have led to a 'global mode of thinking' (20) in which the global links and networks have become more important than a country's human capital. After an analysis of government policies and the use of ICTs in the newly emerging information economies, it concludes that ICT provides for efficiently acquiring, capturing, storing, disseminating, and using local and foreign knowledge on a global basis. This is because of the capacity of ICT to support the development of networks and to establish and maintain connections among individuals, groups, and organizations that possess knowledge considered to be of great use and value to others. In fact, the importance of ICT in supporting knowledge-based development lies in its capacity for efficient networking, interconnectivity, interdependence, and coordination. Whereas physical infrastructure is critical in the industrial age, information infrastructure is becoming indispensable in the knowledge age (25).

The report also introduces the idea of a 'Ubiquitous Network Society' where the 'information can be exchanged any time, anywhere, instantaneously between people, objects, and systems (26)'. It goes on to further look at the economy of this information explosion and the need for Knowledge Based Development by evaluating India as the 'global outsourcing centre. It says, that 'India is well positioned to take advantage of the knowledge revolution to accelerate growth and competitiveness primarily because of the skill and labour endowment of its citizens and its ICT capabilities' (32).

The state-citizen relationship has often been defined significantly by protocols of access to information. The State has been seen as the arbitrator of information dysfunction, and the citizen has been looked at both as the producer as well as the repository of rights and sovereignty over information (with the preamble to the Constitution beginning with the words "We the People"). With the era of satellite telecommunications, another dimension, that of the citizen as a consumer of information entered the discourse. State technocrats like Vikram Sarabhai imagined the contours of participatory democracy as being shaped by people's access to information of the state. The State was hitherto accepted as the collector of citizens' information (via processes like the Census), it was deemed important that the citizens have access to the information (which is often about the citizen) and owned by the State. In this paradigm of information exchange, the State-Citizen relationships have undergone many calibrations as different structures of information production, distribution and arbitration have emerged in the country.

It is not in the scope of this report to produce an exhaustive literature survey of ICT4D and Open data literature. However, the different positions cited here are symptomatic of the kind of questions which have already emerged in the field and which will be addressed in different ways through this report.

This is important for individuals have to feel empowered to use technology which causes them to change the way they interact with institutions and other people. Having ICT infrastructure in place will change people's relationship with technology making it more likely that they will use it to better inform themselves and connect with others.

E-Governance

E-governance initiatives thrive on standardized digital data collection and distribution. The way state and local e-governance initiatives are being created and implemented will determine how easy opening up data will become in the future.

The government has an extensive e-government strategy, suggested in part by the National Knowledge Commission, assembled by the Prime Minister in 2006-2009, charged with making proposals to develop the country's knowledge infrastructure. An e-government strategy was also outlined in the government's 11th Five Year Plan, the country's key general policy planning document, but high level officials at the National Informatics Centre suggests that full implementation of all e-governance initiatives is at least five years and probably a decade away.¹¹ Even if these programs were implemented they do not require opening data in the future.

Even within contexts where ICTs are not the primary object of discourse, the understanding of

¹¹Others, such as Shankar Aggarwal, Joint Secretary and head of the e-Governance Group at DIT, are much more bullish, and think this will happen within the next two-three years.

access remains the same. In a report on Moving Toward Knowledge Based Economies (2007) by the Asian Development Bank, ICTs are considered as one of the four pillars of growth and development in Asia. The report recognises the various forms of e-governance sites which are right now being run in India and looks at them as nodes of open data production and dissemination: “[R]ural Internet kiosks, community e-centers, e-healthcare, geographic information systems (GISs), dairy sector applications, teacher training, online agricultural systems, wireless local loop solutions, databases of rural innovations, and other services targeted at women and children. In the realm of public service, e-Government projects include online delivery of land titles, land and property registration, and empowering dairy farmers through a dairy information and services kiosk”. (33)

These services are being expanded with states like Kerala dedicating a great deal of resources to improving systems and growing infrastructure. However, they are not yet looking at how to make this data available and from several conversations with state officials there is no desire to move toward that end.

Civil Society

There are some civil society groups in India using targeted data in sectors like health, education, elections, and budgets, but there are by no means robust data usage practices among civil society groups. While India has the largest number of NGOs of any country in the world, most are locally oriented and furthermore directed towards specific issues for which there is either no data—as with many tribal issues—or data whose veracity they do not trust, as with school enrolment figures.

Although there certainly are bridges between the government and civil society, it is weak, and it seems that often government officials—at least the higher levels, state and national—are unaware of the civil society groups working in the same space. Perhaps this has to do, again, with the varied and local nature of civil society groups in India, and the lack of truly national issue campaigns—aside from certain things like dalit rights—around which to form cohesive national groups with which the administration must engage. Even national issues tend to be centred in one or two regions; the RTI movement, for example, was rooted firmly in Rajasthan and to a lesser extent Maharashtra.

Moreover, civil society is by and large not very technically literate—probably less so, in many cases, than government institutions themselves, which in most cases have at least a decade of experience with e-governance. Organizations utilizing technology are the exception rather than the rule. Civic hackers, while present, are not very numerous, and it’s unclear to what extent they each are conscious of the work that others are doing, although this could be easily remedied through networking efforts both online and offline. Part of the reason that civic hackers are not numerous is probably that, to this point, there has not been a lot of government data available with which they could engage without first making great efforts to digitize it—for example, often civic hackers working with election data have to produce machine-readable versions of data provided in .pdf or other difficult formats from which to digitally retrieve data. Chethan Elvis, a director with a technology-based social enterprise, believes that once the government opens its data, civic hackers will appear and set to work—so it’s not an issue of willingness or of individual capaci-

ty, but one of limited access to the data on which civil society can innovate. However, Barun Mitra of Empowering India contradicts this and firmly believes that the current problem is a demand-side problem, and that the supply from government will automatically meet demand when it is sufficiently created. Sam Pitroda, adviser to the Prime Minister, also believes that seeing data being put to good use will prove to be a powerful motivator for it to open up even more data. He foresees that “once [government departments] begin to see how data is used, they will pump in more data”.

Media

Reporters Without Borders ranks India at 105 of 175 in its 2009 Press Freedom Index, with 1 being the most free.¹² In general, India enjoys a free press, though the Constitution does not explicitly provide for a free press and the government may restrict speech in certain delimited circumstances. All major newspapers maintain an online presence in various forms and seem well attuned to the uses of technology both for research, news dissemination, and reader interaction—at least with those readers who are themselves wired.

There are many media groups that have gotten themselves involved in a big way in RTI, which is a potent tool for the media. Doordarshan, the national broadcaster which still has the highest penetration in the Indian market and especially the rural market, has a weekly half-hour show on RTI. The show aims to present examples of how common people have used the RTI.¹³ The Doordarshan show travels through the country and present one success story each week from each state. The audience are afforded a chance to call in, and there is also studio-based audience participation.¹⁴

NDTV, a news channel, has a portal dedicated to RTI—information about it, as well as information garnered using it.¹⁵ It even runs an annual award for inspirational RTI activism. Apart from this, it has also covered the physical violence visited upon RTI activists, thus highlighting the need for an effective Whistleblowers Act in India.

The Indian Express newspaper group, along with an NGO named Parivartan, has been instrumental in educating people in exercising their right to information by helping organize awareness camps and training workshops with other organisations. It also runs a regular column called ‘Express Initiatives’. That column publishes sample request to help those who have not ever filed an RTI how simple a process it is, and how to draft one well. It also reports cases of RTI successes and failures. The Indian Express also run an interactive website to address queries and to provide suggestions.¹⁶

Although the proposed Public Information Infrastructure has been covered in the Indian press, most notably in the Economic Times, national news organisations, which seemingly stand to gain greatly from a policy improving access to government information, have not to this point been overly concerned with promoting the idea of open data in India. But B.V. Rao, editor of Govern-

¹²<http://en.rsfsf.org/press-freedom-index-2009,1001.html>

¹³<http://www.rti.org.in/Documents/Case%20Studies/Presentations/RTI%20POWER-Oct-1/Media%20and%20Right%20To%20Information.ppt>

¹⁴Despite this, a 2009 study by PricewaterhouseCoopers found that only 13% of the rural population and 33% of the urban population were aware of the RTI.

¹⁵<http://rti.ndtv.com>

¹⁶<http://expressindia.com/initiatives/rti>

ance Now—a fortnightly magazine and website published by the SAB group, covering the Indian government and focusing especially on issues related to transparency and accountability—believes that journalists are well positioned to take advantage of such a policy once it is implemented successfully. The media is strong even at local levels, he says—both in print and on television—and literacy is improving; access to news is not a problem for most Indians, regardless of where they live.

While Rao says that most media outlets are beholden to corporate interests, and that their actually promoting open data before the fact is doubtful, he says that they have the capacity to utilize open data in their reporting, and furthermore that they will in fact do so. “The corporate sector in India will not do anything other than ask for tax breaks. So I don't see them applying any pressure [for open government data],” he says. But they'd be quick to see its uses,” including the fact that open data allows for everyone to have the same frame of reference. For example, there is currently a debate in India about how many people live in poverty; but there are at least three different figures from different parts of government. Open data would, according to him, go a long way towards enabling stakeholders across the private, civil, and government sectors to make better educated guesses about what the facts on the ground really are. This might be doubtful because not all disagreements on statistics about things like poverty are in fact disagreements over numbers, but sometimes are disagreements over methodology and the way that a particular statistic should be arrived at.

Rao plans to advocate in Governance Now for the Public Information Infrastructure, which he believes will enable public assimilation of data much more than does the RTI Act, because the latter works basically at the individual level and thus creates a false sense that public information is more available than it really is. “RTI has lulled people into complacency because they [mistakenly] think information is available. But only a minor percentage of RTI requests are publicly relevant, he says, and so most government information “is like rainwater in India [because it] goes into the sea and you don't use it.”

Moreover, Rao says, access to what little information is available from the government is sometimes restricted once people actually notice that it is public. For example, Governance Now's online MP Monitor, which provides “report cards” for various members of parliament, consolidates information taken from a government website which, since the media outlet began publicizing the information, has been password protected.

His problems accessing government data aside, Rao is optimistic about the Public Information Infrastructure, and he believes that Pitroda in particular is well positioned to advocate for and implement the changes necessary in government practices for an open government data policy to be a success—“If there's one person who can do it, it's him.” Pitroda is well respected for what he knows, he says, and has the ear of the right people in government ... Nothing in this country moves unless a politician is impressed with a new idea.” As for models on which India might rely in formulating an open data policy—and by extension, the particular form that the PII will eventually take—Rao is not concerned. “When they look at it, they'll look at models [like the U.S. and U.K. initiatives],” he says. “But when the object is openness, what model do you need, anyway?”

Open Data: Why Now?

N. Srinivasan, director of technology transfer and innovation management at the Asian and Pacific Centre for Transfer of Technology of the United Nations' Economic and Social Commission for Asia and the Pacific, has been following e-governance initiatives in India for more than 20 years. Mr. Srinivasan, who is based in New Delhi, identifies three particular historical conditions which together have ushered in an environment in which an open government data policy has become feasible for India, and indeed amenable to its commercial sector.

First, in 1975 the central government started widely using computer systems in its ministries and departments for use in planning and programmes, and in 1976, with assistance from the United Nations Development Program, it created the National Informatics Centre (NIC) to develop and employ e-governance solutions throughout all levels of government.¹⁷ Along with e-governance came automation of processes that before had existed only on paper, and the ability to actually track and manage information in a robust way.

Second, the Right to Information Act of 2005 ushered in “a sea change in India [evidenced by the fact that] the government is willing to is information” to which no one had ever been given access. The RTI Act, he says, was not just a law promoting disclosure of information, but a strategy to enhance delivery of services through a system of citizen-enforced accountability. “The government feels that there's a huge bottleneck between the polices and the delivery mechanisms—what they have developed to reach end beneficiaries—because down the line whole programs get dried up. The delivery rates are very poor.” But the RTI Act empowers individuals to question why they are not receiving benefits, and to collect information on where money and services meant for them are actually going. And, he says, it points the way forward for a national open data policy that is truly proactive in disclosure rather than reactive to individual requests.

Third, as a result of the recent global financial crisis, information technology companies like Infosys and Tata Consultancy have been looking inwards to develop their portfolios, rather than outwards to foreign clients, and these companies and their managers have begun considering in a serious way collaborating with the Indian government as a source of revenue. As an example, he points to the prime minister's 2009 appointment of Nandan Nilekani, a former director at Infosys Technologies, to chair the Unique Identification Authority of India (UIDAI, now National Identification Authority of India). The stage is set, he says—and the financial incentives certainly there—for Indian technology companies to begin developing e-governance and other information products based on open data.

Methodology

Research for this study has consisted of a combination of both primary data collection as well as secondary literature analysis. We have endeavoured to frame the current move towards open data within the evolving relationship in India between the citizen and the government. Primary material has been collected through in-person interviews, phone calls, and e-mail conversations with the important actors in the field, including state and national government officials, transparency activists, “civic hackers”—including both businesses and groups/individuals—and media organizations.

¹⁷<http://home.nic.in/nicportal/aboutus.html>

Policy Environment

Right to Information in India

The idea of open government data — that is, governmental data being made available online, open for bulk downloading and data manipulation — presupposes a general propensity, or at least a general willingness, on the part of the government to provide information to its citizenry. This is by no means a given. Around 82 countries have laws that impose a duty on the government to part with data, usually upon requests being filed by citizens.¹⁸ In many countries, this is a recent phenomenon.¹⁹ In India, the Right to Information (RTI) Act was passed, after a prolonged campaign, in 2005. The campaign for the right to information is, in a sense, highly unusual as it has its origins amongst the rural poor of Rajasthan, and the work of the Mazdoor Kisan Shakti Sangathan (MKSS, literally the Labourer Farmer Empowerment Organization).²⁰ MKSS conceived the right to information as a crucial part of people's audits (of muster rolls, bills, vouchers)—not as a means of finding out that corruption was happening, but rather as a means of officially exposing it, rectifying it, and demanding action against the corrupt. Some of the best accounts of the Right to Information movement and the MKSS struggle are contained in articles by Harsh Mander & Abha Joshi,²¹ Aruna Roy & Nikhil Dey,²² and by Neelabh Mishra.²³

The 2005 Act replaced most existing state-level RTI acts, as well as the Freedom of Information Act, 2002, which was generally acknowledged to be toothless. And, quite importantly, it explicitly states that it overrides the colonial Official Secrets Act, 1923.²⁴ Documents that people have never before had access to, and which the Act specifically notes that the government is not obliged to provide access to, such as minutes of the Union Cabinet meetings, have been revealed in response to RTI requests.²⁵

Shift in Citizen-State Relationship

This shift from a default of secrecy to transparency is a very important one. It not only indicates a concrete duty on the state to provide information as per the law, but can also be seen as an indication of a shift in the very conceptualization of the body politic. Formerly passive subjects

¹⁸ Roger Vleugels, Overview of all 86 FOIA Countries, at <http://www.right2info.org/laws/Vleugels-Overview-86-FOIA-Countries-9.08.pdf>

¹⁹ The majority of the countries with such laws have adopted them post-2000. *Ibid.*

²⁰ States like Goa and Tamil Nadu did bring about their state-level Right to Information Acts without the push of a grassroots movements. However, even in those states, the national-level campaign started by the MKSS and carried forward by the National Campaign for People's Right to Information did have an effect in creating a space to enforce accountability.

²¹ Harsh Mander and Abha Joshi, Movement for Right to Information in India: People's Power and Control of Corruption, available at <http://www.humanrightsinitiative.org/programs/ai/rti/india/articles/The%20Movement%20for%20RTI%20in%20India.pdf>

²² Aruna Roy and Nikhil Dey, The Right to Information: Facilitating People's Participation and State Accountability, available at <http://www.10iacc.org/download/workshops/cs54b.pdf>

²³ Neelabh Mishra, People's Right to Information Movement: Lessons from Rajasthan, available at <http://data.undp.org.in/hdrc/dis-srs/Rajasthan/Right%20to%20Info.pdf>.

²⁴ Section 22 of the RTI Act states: "The provisions of this Act shall have effect notwithstanding anything inconsistent therewith contained in the Official Secrets Act, 1923, and any other law for the time being in force or in any instrument having effect by virtue of any law other than this Act."

²⁵ Minutes of Cabinet meetings may now be found on the website of the Commonwealth Human Rights Initiative.

of a state are re-imagined as active citizens with a legitimate interest in such information, and thus having an inherent right to it.²⁶

Indeed, the seed of the call for 'open government data' can be found in the Right to Information Act itself. Section 4 of the Act states: "It shall be a constant endeavour of every public authority to take steps ... to provide as much information *suo motu* to the public at regular intervals through various means of communications, including Internet, so that the public have minimum resort to the use of this Act to obtain information."

This in a way encapsulates the rallying call for open government data advocates. It goes one step beyond the concerns of Western open data advocates, in that it does not rely upon the Internet as the sole medium of communication. As pointed out by Venkatesh Nayak of Commonwealth Human Rights Initiative, sometimes a blackboard outside a panchayat office is the most effective mode of communication. However, the RTI Act does not address the concerns that they have about the reusability of the data, which is governed by copyright law.

Sruti Bandyopadhyay of the Centre for Policy Research looks at the RTI act as an indicator of a shift in official attitudes from that of the Official Secrets Act. She points out that public consultations regarding legislations is much more prominent and frequent now. Even the broadcast of the proceedings of the Lok Sabha and the Rajya Sabha through dedicated free-to-air channels are markers of such a shift, she notes.

"Information" and Raw Data

The RTI Act provides for a very wide definition of the word "information" as "any material in any form, including records, documents, memos, e-mails, opinions, advices, press releases, circulars, orders, logbooks, contracts, reports, papers, samples, models, data material held in any electronic form and information relating to any private body which can be accessed by a public authority under any other law for the time being in force". As this is an inclusive definition, it actually covers "material in any form" apart from those explicitly mentioned as well. The raw data that goes into reports will also get covered by the word "information" under the RTI Act.

Criticisms of the RTI Act and Processes

According to the 2009 Global Integrity report, India scores an 85 on "Public Access to Information" earning itself a rating of "strong". However, ratings hide as much as they convey. There are many shortcomings with the implementation of the RTI Act. There have been many surveys on this subject,²⁷ so this section shall try to concentrate on only a few of the shortcomings, in-

²⁶ The preamble to the act reads:

WHEREAS the Constitution of India has established democratic Republic;

AND WHEREAS democracy requires an informed citizenry and transparency of information which are vital to its functioning and also to contain corruption and to hold Governments and their instrumentalities accountable to the governed;

AND WHEREAS revelation of information in actual practice is likely to conflict with other public interests including efficient operations of the Governments, optimum use of limited fiscal resources and the preservation of confidentiality of sensitive information;

AND WHEREAS it is necessary to harmonise these conflicting interests while preserving the paramountcy of the democratic ideal;

NOW, THEREFORE, it is expedient to provide for furnishing certain information to citizens who desire to have it.

²⁷ See, e.g., Global Integrity Report Card on Indian Public Access to Information:

stead of aiming to be exhaustive.

One of the most glaring procedural shortcomings of the Act is that the process for appointment of Information Commissioners is still very opaque. As the Information Commissioners are crucial part of the system as the people who proactively enforce the RTI Act and are also the appellate authorities for unsatisfied RTI applicants. Shailesh Gandhi, an Information Commissioner with the Central Information Commission and formerly an RTI activist, criticised his own appointment, noting that it too was non-transparent.

Additionally, in a study conducted by PricewaterhouseCoopers, 75 per cent of the respondents noted their dissatisfaction with the information furnished by the public authorities.²⁸ It is often the case that incomplete or irrelevant information is provided. Very often, it takes more than the stipulated time period of 30 days to receive the information. This is usually due to the poor record-keeping within the public authorities, and is a more fundamental problem of a sorely lacking information architecture. This is one of the crucial factors in the non-compliance of public authorities with s.4(1)(b) of the RTI Act, which requires proactive disclosure.

The proactive disclosure requirement of the RTI Act has not met with much success. Research independently done by India Governs, CHRI, and others confirms this.²⁹ As noted above, one of the largest problems with complying with the proactive disclosure requirements of the RTI Act is that there is no easy system through which this data can be published online. The difficulties highlighted in the 'Challenges' section below are amongst the reasons for the failure of implementation of the proactive disclosure requirement. Shankar Aggarwal, Joint Secretary (e-Governance) with the Department of IT, says that the issue with Section 4 is "whether ... the ministries are implementing the provisions of the act in letter and in spirit or ... only in a perfunctory manner. That is the only issue. So ... the nodal [RTI] agency has to issue some kind of direction to all the ministries and depots to ensure that all the data which is being kept by that particular department is available to the public ... in an efficient manner." Such a directive, he says, should be the concern of Department of Personnel and Training, which is responsible for implementing the RTI Act.

Many officials also complain that while the RTI Act invests duties on those designated as Public Information Officers, those duties are usually in addition to their regular duties and without any additional help being provided for information retrieval, and require working with poor information systems. This, coupled with penalties for renegeing on their duties, is a source of legitimate complaint for PIOs. "Frivolous RTI requests" are also a reality, though exaggerated,³⁰ and the measures the government's proposed to curb them are universally condemned by civil society.

One other complaint that the transparency-focussed laws (such as the RTI, and the mandate on

<http://report.globalintegrity.org/India/2009/scorecard/15>

²⁸ http://rti.gov.in/rticorner/studybypwc/key_issues.pdf (section 4.2.3, "Poor Quality of Information Provided")

²⁹ The Report of People's RTI Assessment 2008 notes: "Despite a very strong provision for proactive (suo moto) disclosure under section 4 of the RTI Act, there is poor compliance by public authorities, thereby forcing applicants to file applications for information that should be available to them proactively, and consequently creating extra work for themselves and for information commissions." Available at http://www.rti-assessment.org/exe_summ_report.pdf

³⁰ Himanshi Dhawan, RTI officials say no misuse of act, only frivolous requests Read more: RTI officials say no misuse of act, only frivolous requests, The Times of India, 13 Sep 2010, available at <http://timesofindia.indiatimes.com/india/RTI-officials-say-no-misuse-of-act-only-frivolous-requests/articleshow/6548574.cms>

electoral candidates to provide information on assets) is that while they promotes transparency, they do little to convert that into accountability. Thus, while assets of an electoral candidate might be declared and made public, those are rarely matched up with their income tax returns, and questions of ill-gotten wealth remain questions even when it is plain for everyone to see.

Shortcomings of Reactive Disclosures

Though proactive disclosure faces many problems, many argue that official workload would decrease if proactive disclosure is followed. Many RTI requests are just repeat requests for information that has previously been granted. If proactive disclosures become the norm, and responses to RTI requests are carefully archived on each public authority's website and made easily available, there is a good chance that the number of requests might go down. This is very important given that most Information Commissions already have large pendencies—while CIC has a pendency of almost a year, Maharashtra and UP have pendencies of almost two years. Some others, however, are skeptical on this claim about reduction in workload, as they feel most of the requests are highly individualised, and not requests for standard data.

There is evidence that proactive disclosure will definitely be seized upon by many in civil society organizations and technology firms. Most of the case studies and the groups that are surveyed in the Appendix to this report demonstrate that. As briefly noted earlier, Chethan Elvis, a director at Mahiti—a social enterprise focussing on free/open source software-based technological solutions for both the non-profit and commercial sectors, with an emphasis on social causes—says that organizations like his would likely begin innovating on top of open government data were it to be released in machine-readable formats. Although Mahiti's work is driven by the needs of its customers, he believes that many would see the value of working with such data and commission projects making use of it.

Similarly, Krishnaraj Rao, an RTI activist and former journalist, believes that civil society has the capacity to use open government data to hold officials accountable, as long as that data is available. He believes that there is already a network of RTI stakeholders "who are willing to analyze, process the data, understand its context, and put it in front of the press and the people ... so that it can be understood." This network, he says, can be well utilized in any open government data scheme.

One somewhat radical idea proposed on this issue by Shailesh Gandhi—an RTI Activist turned Information Commisisoner at the Central Information Commission—is that of flipping the workflow. Currently, public authorities decide what they wish to reveal to the public (usually synonymous with what they are required to reveal to the public), and build their workflows around that. Instead, he proposes that workflows should be based on deciding what information will not be put up, and making available everything else online. Information overload, he believes, is not a problem that public authorities should concern themselves with, and believes that the public can deal with it well enough, especially given modern technology.

Government Copyright in Law

Just as in other countries³¹ not much attention is normally paid to government copyright and use of government-produced materials (to be clear, not government-funded copyrighted works, which is a much larger category). In India, the government is the owner of copyright over all works produced by government employees (by virtue of section 17(d) of the Copyright Act, read with section 2(k)) and of all materials produced by any Indian legislative or judicial body. Government copyright lasts for 60 years from the date of creation of the work (per section 28 of the Act). There is a limited exception to government copyright in the form of section 52(1)(q):

52(1) The following acts shall not constitute an infringement of copyright, namely: (q) the reproduction or publication of — (i) any matter which has been published in any Official Gazette except an Act of a Legislature; (ii) any Act of a Legislature subject to the condition that such Act is reproduced or published together with any commentary thereon or any other original matter; (iii) the report of any committee, commission, council, board or other like body appointed by the Government if such report has been laid on the Table of the Legislature, unless the reproduction or publication of such report is prohibited by the Government; (iv) any judgement or order of a court, tribunal or other judicial authority, unless the reproduction or publication of such judgment or order is prohibited by the court, the tribunal or other judicial authority, as the case may be.

Government Copyright in Practice

While the exception in the law may seem narrow, there are very few cases of the government actually asserting its copyright against any individual or group, even when the publishing of the government work has been a commercial endeavour. A former Registrar of Copyright noted how in the 1990s there were cases of government-commissioned reports being sold for profit, but that the Secretary of the department which had commissioned the report shrugged it off. The present Registrar of Copyright, and Deputy Secretary in the Department of Secondary Education, Mr. G.R. Raghavendar noted how he couldn't recall a single case of the government pursuing a case of infringement since he took over his present post. However, he did note that a case might be instituted against pirates by the governmental body NCERT which publishes school textbooks and sells them at a subsidised rate (and has also made them available gratis online).³²

Even in cases where information is made publicly available, such as on the India Water Portal, on the website of the Reserve Bank of India, and others, copyright notices are still to be found, and no thought has been given to the licences.

³¹The losses suffered by NCERT due to such piracy is questionable. From Mr. Raghavendar's account itself, NCERT loses money with many of the books it sells since it sells them at a subsidised rate (books up to Class 7). It pays authors *despite* publishing subsidised books, and not out of the profits generated from publishing books. Further, the pirates are apparently selling the books at the same rate as NCERT or even cheaper, so the students are not put to any loss either and might even gain.

³²The losses suffered by NCERT due to such piracy is questionable. From Mr. Raghavendar's account itself, NCERT loses money with each book it sells since it sells them at a subsidised rate. It pays authors *despite* publishing books, and not out of the profits generated from publishing books. Further, the pirates are apparently selling the books at the same rate as NCERT, so the students are not put to any loss either.

Copyright and Right to Information

Talking to RTI activists, it emerged that there have been instances where the authority providing information under a right to information request has asserted that the information provided was not to be shared with anyone else.³³ However, as far as is ascertainable, in none of those cases has copyright been cited as a reason for not sharing information.³⁴ Mr. Raghavender asserted that all RTI responses are copyrighted, and that is probably the correct understanding of the law, since the RTI Act itself does not seem to exempt responses from copyright law. But as noted above, the law does not always match up to practice, and that is a good thing. Some activists, such as Mandakini Devasher of Accountability Initiative, disagree with Mr. Raghavender's reading of the law, and believe that by virtue of sections 8 and 9 of the Act the public information that is provided is rightfully in the public domain, and out of copyright.

Copyright and Data

It must be remembered that government data are usually just figures and statistics. It is questionable whether such data would be considered 'original' literary works, as non-original literary works are not protected by copyright law. In a landmark case in 2007, the Supreme Court of India changed the standard of originality from one based on labour-expended (the "sweat of the brow" doctrine) to one based on creativity, skill and judgment. The bulk of raw government data would probably be adjudged to have involved much work to gather, but insufficient creativity to merit copyright protection. Notably, while copyright extends to original databases in India, it does not extend to non-original databases. The idea of database rights (giving exclusive rights over non-original databases), recognized in some other parts of the world, are not recognized in India.

Standards and Interoperability

E-Government Interoperability Framework

Electronic data is stored in a multitude of formats. Most of these formats are mutually undecipherable. Thus, one government department saving its documents in a scanned format as TIFF files cannot allow for other departments to interact efficiently with those files. For such reasons, certain set standards should be used by all government departments. Of the different standards that can be used, standards that have been developed in an inclusive and participatory fashion and which is available freely for implementation ought to be preferred. Such a standard, very importantly, prevents vendor lock-in, as many vendors (including the government itself) are amenable to implement their own open standard.

³³One notably public instance of this was when the Padma awards nominees list was disclosed by the Home Ministry following an RTI request by activist S.C. Agrawal, but the reply advised Mr. Agrawal that the information was for his "personal consumption" only, and asked him not to disclose it to the media; he promptly sent over the reply to a number of newspapers. See Utkarsh Anand, *Activists Ask Why Rider on Padma Award RTI Reply*, 09 Apr 2010, available at <http://www.expressindia.com/latest-news/Activists-ask-why-rider-on-Padma-award-RTI-reply/602085/>.

³⁴While reacting to the story on the Padma awards, Chief Information Commissioner Wajahat Habibullah said a public authority could only "recommend" such requests but cannot impose such terms. The Indian Express article quoted him as saying: "The Act speaks only about disclosure or non-disclosure of the requisite information and no public authority can press a condition in their reply. Therefore, though it could be a matter of courtesy, there cannot be any legal obligation on an applicant to abide by any such term." *Id.*

Open Standards Policy

Recognizing this, the Department of IT released a draft National Policy on Open Standards in August 2008 and invited responses. Eventually, in May 2009, a more recent draft of the policy was leaked, and publicly made available by Fosscomm. This draft diluted the requirements to classify as 'open' standards quite considerably, allegedly on the insistence of the software industry body NASSCOM. This resulted in a public outcry with many people posting about this controversy, and writing in to the DIT. Finally in May 2010, an official second draft was released for public comment.

When the Joint Secretary (E-Governance), Mr. Shankar Aggarwal, was interviewed for this report on August 3, 2010, he stated that the policy would be notified "soon". A chief problem in ensuring the use of open standards is that every government ministry has an independent IT section providing information and services, to which the NIC may provide help on a consultancy basis. "The main point is that there is no one policy for the entire country. Many countries have come up with one policy saying that even though we do not have this capability right now because, the information is not digitised, this is where we want to be headed," Dr. Govind notes. "But in India all of this depends on every single department, and many of them aren't really thinking about these issues."

Machine Readability

Not all open standards (such as PDF) are 'machine-readable' so that the data can be manipulated, reprocessed, visualized, mashed-up with other data, or even made interactive. While it is desirable to have information put up in open standards, it is also desirable for them to be in machine-readable formats (such as well-defined XML).

There is a case to be made out for data to be put out in both a human-readable as well as a machine-readable format. The former is an important part of ensuring that non-technical citizens have as much access to the data as more technically-inclined citizens and technically-capable organizations and groups. Apart from technically-capable organizations and groups, machine readability enhances search, and is an important component of electronic accessibility.

A working definition of a machine-readable format could be: "A format that has been designed to be processed and manipulated by computers without the requirement of human parsing."

However, even machine-readability does not guarantee ease of use for civic hackers. For instance, text PDFs (as opposed to image PDFs) and even plain HTML files are machine readable, but most often are not as ideal as well-structured XML since the context and the metadata would be provided in the well-structured XML file, while they would have to be extracted from the PDF and the HTML.

Electronic Accessibility Policy

There are many groups in society who are unable to access information and services which are made available over the Internet, including people who are disabled, uneducated or elderly. There are some universally recognized standards and guidelines which provide the means to alleviate such as the Web Content Accessibility Guidelines (WCAG 2.0) of the World Wide Web Consortium

(W3C), which many website developers are not aware of, and which a large percentage of websites are not compliant with.

Recognising that adoption and adherence to accessibility standards was vital for ensuring participation of the 70 million persons with disabilities living in India, as well as millions of elderly and illiterate persons, the Department of Information Technology (DIT) under the Ministry of ICT has taken the initiative to formulate a national policy on electronic accessibility, which would apply to all government funded infrastructure across the country. The process has been transparent and participatory with deep involvement and collaboration from both disability groups and industry, and provides an example of best practices that may shed light on how best to coordinate the formulation of a robust national open data policy in India.

The policy process began with a national stakeholder consultation where participants from the NIC, central government departments, civil society groups such as the Centre for Internet and Society, independent consultants like Barrier Break Technologies, industry representatives like Microsoft, and research organisations like the Centre for Development of Advanced Computing achieved consensus on the need to formulate a policy on Internet and electronic accessibility. After many rounds of consultations and research, a draft was prepared and put up for public review for two months. Finally, the DIT consolidated the final draft, which it sent to all central ministries and state departments for feedback, and which it is still awaiting.

As pointed out a bit earlier, all accessible documents on the web are also machine readable and searchable. Thus ensuring strong electronic and web accessibility policies and their strict enforcement on government websites results, automatically, in machine readability.

National e-Governance Plan

In May 2006, the Indian government approved the National E-Governance Plan (NeGP), which was conceptualized as a holistic approach towards making government services available to people in their localities through CSCs while meeting goals of efficiency, transparency, reliability, and affordability. In short, the plan's "vision is to use Information Technology as a tool for raising the living standards of the common man and enriching their lives."³⁵ The plan includes proposals for "streamlining, aligning, optimizing and automating all internal processes across government boundaries"; with respect to courts, "online availability of judgments and cause list, e-filing of cases and notifications through e-mails"; and a portal providing "one-stop access to government services."³⁶ The NeGP also lays the groundwork for the UID as well as statewide area networks and data centers, and calls for research into "e-Government Enterprise Architecture Frameworks, Information Security, Data and Metadata Standards," among other areas.³⁷ Most importantly, probably, the plan calls for "establishing 100,000 broadband Internet enabled Common Service Centers (CSCs) in rural areas of the country."³⁸

Figure 1.

Mission Mode Projects in India

³⁵Chauhan, Radha. "National E-Governance Plan in India"

<http://www.iist.unu.edu/www/docs/techreports/reports/report414.pdf>, 6.

³⁶Ibid., 7.

³⁷Ibid., 12.

³⁸Ibid., 10.

Central MMPs	State MMPs	Integrated MMPs
Banking Central Excise & Customs Income Tax (IT) Insurance MCA21 National Citizen Database Passport Immigration, Visa and For- eigners Registration& Track- ing Pension e-Office	Agriculture Commercial Taxes e-District Employment Exchange Land Records Municipalities Gram Panchayats Police Road Transport Treasuries	CSC e-Biz e-Courts e-Procurement EDI For eTrade National e-Governance Ser- vice Delivery Gateway India Portal

Common Service Centres

In 2004, the government announced a nationwide initiative to establish 100,000 Common Service Centres (CSCs) serving 600,000 villages with the objective of developing a platform enabling government, private sector, and civil society organizations “to align their social and commercial goals for the benefit of the rural population in the remotest corners of the country through a combination of IT-based as well as non-IT-based services.”³⁹

Under the auspices of the Department of Information Technology, the CSCs are core components of India's National e-Governance Plan and are positioned as “change agents” to promote “rural entrepreneurship, rural capacities and livelihoods, enable community participation and effect collective action for social change - through a bottom-up model that focuses on the rural citizen.”⁴⁰

The CSCs, each of which is designed to serve a cluster of six to seven villages, are designed as ICT-enabled kiosks containing PCs and basic support equipment like printers, scanners and are each linked to a national data network. They are also manned by staff who can assist illiterate citizens—either with respect to reading or using technology—in utilizing their services, similar to how Indian postal workers used to read letters to recipients. As of July 2010, about 83,500 CSCs had been established throughout the country.⁴¹ These kiosks will undoubtedly play an important role in any open data policy seeking actually to provide information to citizens otherwise lacking access to data networks.

Importantly, too, many in government understand the CSCs as not only providing services, but also empowering citizens who otherwise would not be able to contribute significantly to Indian society. There is from the beginning a political component to enabling access to information and services at the village level.

“We have to empower these people...who are living in villages. Once they get empowered, even if they contribute a little bit, it will be an immense contribution,” Mr. Shankar Aggarwal says. “Today people who are living in villages do not have good quality education, do not have good

³⁹<http://www.indg.in/e-governance/cscscheme/common-service-centres-scheme>

⁴⁰<http://www.csc-india.org/>

quality health services, do not have bank accounts. With the help of ICT all this can be achieved...[I]n another two or three years there will be a paradigm shift in the way we do our business, and people at the grassroot level will get empowered and will join the mainstream. There is no doubt. Nobody can stop this.”

Associated closely with the idea that people must be empowered at the village level is the theory that when those at the lowest levels of society are empowered, so will be India as a whole. “Today if you [look at] the [Indian Administrative Service], most of the guys are coming from smaller towns. If you go to [Indian Institutes of Management] and [Indian Institutes of Technology], most of [the students] are from the middle classes. Most of them are from smaller towns or even villages,” Mr. Aggarwal says. “What does it show? That people are getting empowered. And once they get empowered then the masses will get empowered and the entire nation will get empowered.”

The key point, perhaps, is that whatever side of increasing the provision and dissemination of data one finds herself on—whether as a bureaucrat or as a citizen, as an urbanite or a village dweller—the political stakes from the outset are already defined as very high.

National Knowledge Commission Recommendations on e-Governance

In June 2005, Prime Minister Manmohan Singh constituted the National Knowledge Commission, an advisory body to the Office of the Prime Minister, (NKC) with the mandate to recommend policy reforms in the areas of “access to knowledge, creation and preservation of knowledge systems, [and] dissemination of knowledge and better knowledge services.”⁴² The Commission was chaired by Sam Pitroda, a popular national figure known for reforming the country’s telecommunications systems who in 2009 was appointed to a cabinet-level position as Adviser to the Prime Minister for Public Information Infrastructure and Innovations, and in August 2010 was named chairman of the newly formed National Innovation Council.⁴³ The NKC was given a period of three years to conduct research and develop recommendations, which it issued in a series of reports now compiled in the “National Knowledge Commission Final Report 2006-2009.”⁴⁴ Its recommendations on e-governance were prepared by a committee chaired by Nandan Nilekani.

In its Final Report, the NKC, asserting that “Providing access to knowledge is the most fundamental way of increasing the opportunities of individuals and groups” (13), made two recommendations particularly relevant to implementing an open government data in India. First, the NKC “recommended the establishment of a high-end National Knowledge Network connecting all ... knowledge institutions in various fields and at various locations throughout the country, through an electronic digital broadband network with gigabit capacity” (13). Second, and more relevant to considerations for open government data specifically, the NKC proposed that the government create a series of “national web based portals on certain key sectors such as Water, Energy, Environment, Teachers, Biodiversity, Health, Agriculture, Employment, Citizens Rights etc. [serving] as a single window for information on the given sector for all stakeholders and ... managed by a consortium consisting of representatives from a wide range of stakeholders” (13).

⁴²<http://www.knowledgecommission.gov.in/about/default.asp>

⁴³<http://sify.com/news/pm-sets-up-national-innovation-council-news-national-kiquEfdecge.html>

⁴⁴<http://www.knowledgecommission.gov.in/reports/report09.asp>

Following the NKC's recommendations, the government facilitated public-private partnerships establishing five portals intended to "become a decisive tool in the popular movements in support of the right to information, decentralisation, transparency, accountability and people's participation [and] to increase openness and enhance accessibility" (39), the India Water Portal, the India Energy Portal, the India Environment Portal, the India Biodiversity Portal, and the Teachers of India Portal.

The NKC explicitly recommended that "[a]ll government departments should easily make available data sets they have, in a digital format to the portal consortium." It is unclear to what extent this recommendation has been followed; currently, portals are mostly hosting information provided by NGOs, research and academic organizations, and individual users. The NKC recognized that "data that is traditionally collected and managed separately, unrelated to each other, should now be seen together. But it indicated that "[t]here are no platforms or mechanisms currently in place to allow this to be done easily" and recommended also the development of clear guidelines for appropriate data formats as well as the regular updating of hosted data, suggesting that the RTI Act could play a role in fostering simpler rules (39-40).

Additionally, the NKC recommended that portal teams "work proactively with NGO and Government networks, use mass distributions channels like radio, television and the print media to ensure ... knowledge is leveraged to precipitate change on the ground" as well as develop non-Internet-reliant software for data dissemination that could both download and upload data at central networking location. Importantly, too, portals must seek to provide data in local languages (40). These recommendations would be well taken into account in the implementation of any serious national open data policy, and likely will be considering that the chairman of the NKC, Mr. Pitroda, is responsible also for developing India's public information infrastructure.

Public Information Infrastructure

In 2009, Prime Minister Manmohan Singh appointed Sam Pitroda to the cabinet-level position of Adviser to the Prime Minister for Public Information Infrastructure and Innovations, tasked with developing a unified policy for information standards and practices incorporating both intra-government affairs and citizens' services.

Mr. Pitroda, who divides his time between Chicago and New Delhi, has enjoyed a long association with the Nehru-Gandhi dynasty, India's leading political family, since Indira Gandhi was prime minister. Mr. Pitroda served as technology adviser to Prime Minister Rajiv Gandhi, and works closely today with Rahul Gandhi, who is currently an MP and general secretary of the Indian National Congress Party—and often mentioned as a future prime minister. In particular, Mr. Pitroda and Mr. Gandhi have recently traveled together to villages introducing broadband services. Mr. Gandhi's mother, Sonia Gandhi, is currently president of the Indian National Congress Party, of which the current prime minister, Mr. Singh, is the legislative leader. Mr. Pitroda is also a popular figure among the public, having played a major role in transforming India's telecommunications infrastructure in the 1980s and 1990s. He is well known, well connected, and as well positioned as anyone in the country to advocate for sweeping policy changes at all levels of government.

In June 2010, Mr. Pitroda's office uploaded online a slide presentation on "Strengthening De-

mocracy and Governance: Public Information Infrastructure."⁴⁵ The presentation provides a basic overview of his proposal for a robust information system implicating all levels of government but focusing access and delivery on the level of the panchayat, or village assembly, which it specifies as the nodal point for citizen services.

The presentation declares information to be a "public good" and envisions that the PII will, among other goals, "radicalize democracy through informed citizen participation[,] improve delivery of services[,] empower local governance & community institutions[,] and enhance equity and efficiency." Additionally, the PII seeks explicitly to "radically transform governance" by providing "new opportunities for 'crowd'-sourcing of ideas, feedback and evaluation from citizens on governance issues." The PII is the closest scheme yet proposed in India to a portal like data.gov or data.gov.uk, and in many ways it goes beyond either.

The PII consists basically of five elements: CSCs; a core backbone consisting of the National Knowledge Network, connecting 1,500 institutions across the country with gigabit capabilities; data centers, including 35 state and four national centers, which will also be used in the UID program; a security framework; and applications and platforms enabling people to access information as well as analyze and innovate upon it.

Included in the scheme is a national repository of information on people, including citizenship, resident, and household data; places, including villages, towns, streets, schools, hospitals, government offices, factories, officers, residences, stations, mines, minerals, dams, plants, rivers, parks, forests, farms, etc.; and programs and other government offices, such as the National Rural Employment Guarantee Scheme, the Public Distribution System, girl child benefit schemes, pensions, the judiciary, police and prisons, treasuries, land records, universalization of elementary education, and the National Rural Health mission, among others. Furthermore, the PII will incorporate the UID project as "the basis of schemes and programs of central and state programs and for identifying rights beneficiaries."

Applications hosted on the PII will include a shared Geographic Information System (GIS) for the Survey of India; the National Disaster Management program; the Urban Ministry; the Departments of Space, Security, Environment, Health, and Rural Development; the Planning Commission; as well as private enterprises. Data from these entities will be publicly available on a single portal accessible by a variety of clients, including PCs and mobile phones. The portal will also incorporate applications, communities, mash-ups, and allow for a variety of analyses on data including including survey, remote sensing data, census, education, and health data, as well as forest, land use and groundwater data.

In order to meet its goal of serving citizens at the panchayat level, the PII will seek to provide broadband connectivity to about 250,000 panchayats and will focus primarily on enabling public access to education, health, employment, productivity, and disaster relief, as well as providing commercial services like banking, rural ATMs, wholesale and retail procurement, secured transactions, and entertainment. In order to facilitate delivery of these services, the PII will seek to establish paperless environments in government officers by fully automating government processes via electronic work flows and file management systems.

⁴⁵<http://www.slideshare.net/pmpiii/public-information-infrastructure-4560021>

While this might seem like an impossible task, many believe it can be done. B.K. Gairola, Director General of the National Informatics Centre, points out by the end of the year, India will have set up a national optical fibre network. Once the country has actually decided to do something, he says, it moves quickly, and the PII will be no different. He has full faith in India's ability to develop the necessary technology for the PII and implement it throughout the country, as well as in its capacity to educate citizens about how to use the data and other IT tools with which the PII will provide them.

Mr. Pitroda believes that India shouldn't go to a company like Oracle or IBM to develop the technology that it needs to implement the PII, but rather should develop it locally. Further, he thinks India faces a distinct advantage while doing so. "What we are trying to do has never been done," he noted; India's advantage is that, unlike the U.S. and the U.K., it is not reliant on legacy technology. Rather, it can "leapfrog" initiatives like data.gov and data.gov.uk to create a new kind of information infrastructure.

And perhaps most importantly, Mr. Pitroda firmly believes that the Government is ready to open up the system—that there is political will currently to push forward with greater transparency

National Data Sharing and Accessibility Plan (NDSAP)

The National Data Sharing and Accessibility Policy (NDSAP)⁴⁶ would setup a framework that would create a DATA.GOV.IN. It is designed to apply to all non-classified data that is publicly held by various government departments. "The NDSAP policy would help data users and data solicitors get access to data through established procedures and defined norms."

The objective of the policy *"to address all issues related to data in terms of the available scope of sharing and accessing spatial and non-spatial data under broad frameworks of standards and interoperability. "a) Data Classification; b) Technology for sharing and access; c) Current legal framework (RTI Act and Privacy Act)"*

It includes India's own version of the 10 principles of open data discussed above *"Openness, Flexibility, Transparency, Legal conformity, Protection of intellectual property, Formal responsibility, Professionalism, Interoperability, Quality, Security, Efficiency, Accountability, Sustainability."* There is very little overlap between the 13 these open data principles and the 10 principles of open data examined above. The law also offers no explanation for these terms or how they will affect data. The biggest difference is that the Indian principles keep the data propriety while the international versions stress the opposite.

What stands out in the policy is that the departments have to create a list of data that should not be shared which implies a default openness stance. Each department will provide a list of un-shareable items that will be determined using the provisions in the RTI Act and Privacy Act. Then all other data sets will be considered safe to be open to the public.

MetaData would also be provided which would allow people to know what data is available. A three pronged classification system would be created to deal with different types of data; Open Access, Registered Access, and Restricted. A data warehouse will be set up to house current and historical data so that this information in is one place.

⁴⁶ <http://www.dst.gov.in/nsdi.html>

The NSDI, which has written the policy, under the Department of Science and Technology solicited feedback for the policy of which Center for Internet and Society submitted comments which includes changing the categorization and requiring the data be in open standards.⁴⁷

It is still unclear the future of this policy, what the next steps are and what department will be in charge of Data.gov.in. But if it goes through it would give data policy roots and government information a central place to be found and evaluated.

Case Studies

While the policies surrounding open government data and e-governance look promising, implementation is another matter and real world experience must be examined to fully understand what works and what doesn't.

One of the things found by researchers throughout the sphere is that the term "open data" was understood generally to refer to the proactive disclosure of information. The idea that information should be free to use, reuse, and redistribute, and that it should be available in open and machine-readable formats, was not so much rejected as it was set aside in favor of maintaining that data is "out there," that "everything is online." The assumption seemed to be that the mere fact of data or information was enough, and that the particular mode in which it appeared was of secondary importance if any.

The following case studies are structured to address what government is doing around data and e-governance and their attitude toward open data. CSOs who are actively filling in gaps in government data and analysis; organizations working with government to change attitudes and organizations using data to impact communities. All while looking at the issues that are being encountered by these different experiences.

Government Case Studies

The following experiences are from government e-governance projects from the national level and the states of Maharashtra, Tamil Nadu and Kerala. These projects are dealing with issues around data collection, data standardization, ICT infrastructure building and access.

Maharashtra Department of Agriculture

Although Maharashtra is India's leading industrial and one of its most developed states, a majority of its citizens are engaged in agricultural work. Several of its departments, including the two covered here—Agriculture and Environment—have been using ICT for more than two decades, and their respective ICT practices represent some of the best of state governance while illustrating the limits of and problems with such practices in the present.

The Maharashtra Department of Agriculture, located in Pune about 170 km from the state capital of Mumbai, has been using ICT since 1986, when it implemented a computerized system to process census data. The department was mentioned by several sources as being widely acknowledged among officials as having a robust record-keeping policy. Currently, the department employs ICT both for internal administrative matters and to collect, maintain, and disseminate infor-

⁴⁷Centre for Internet and Society 'Comments on NDSAP' (<http://cis-india.org/advocacy/openness/blog/draft-ndsap-comments>)

mation to farmers across the state, both online and via SMS. The department maintains a mostly bilingual website⁴⁸ in English and Marathi, the local language of Maharashtra.

The department's website currently hosts data and statistics related to the participation of Maharashtra farmers in the National Agriculture Insurance Scheme; annual growth rates of agriculture and animal husbandry; rainfall recording and analysis; and crop reports by various state administrative divisions, crop type, and related pests and diseases. The website also features soil and crop, horticultural, soil/water conservation, agricultural inputs, statistical, and district-wise fertility maps. Additionally, the website allows farmers and other interested parties to sign up for a crop-specific SMS service advisory, available in both English and Marathi, subject to the ability of the end users' handsets to render Marathi. This year, the department has offered a service advisory on 43 different crops, for which it has sent about 26,000 SMS messages to about 40,000 farmers. As of June 25th the site has received 3,069,108 hits in total.

Balasaheb Thorat, commissioner of agriculture for Maharashtra, says that his department employs ICT in three areas: daily administrative business, information and knowledge dissemination to stakeholders, and delivery of services. The problem that he faces using ICT, though, is that relatively few people have access to the necessary technology to take advantage of his offerings both online and via SMS. There is, first of all, the question of whether people have electricity, and second whether they have Internet and/or mobile phone access—most have the latter, but not the former. After the question of access comes that of willingness and capacity; many farmers, he says, are still stuck in a mindset requiring department workers to travel to their villages and provide people with information directly. Even where farmers do have Internet and/or mobile access, they either do not know how to, or do not want to, request or download information provided by his department. "They require someone in person to attend to them," he says.

Notwithstanding those farmers who lack access or otherwise fail to get information from his department on their own via ICT distribution routes, he says, there is a group of farmers that are comfortable using ICT and employ it to track everything from weather to pests to international agricultural markets. And he sees a trend developing whereby more and more farmers will begin using ICT as services become, he says, ever more transparent, effective, efficient, and timely, especially as the department—and other government offices—develop applications for online services, including one in particular where farmers can apply online for government benefits. The problem of access, he thinks, will be solved by increased penetration of and citizens' familiarity with common service centres, of which there are already about 10,000 spread throughout Maharashtra.

While utilizing ICT structures and building infrastructure and partaking in access solutions there is a lack of attention to opening the data created by this system. The department's website features a wealth of information, it is only available — including graphs and spreadsheets — either in PDF, JPG, or HTML formats from which information either cannot be easily extracted (in the case of .pdf and .jpg files) or downloaded (in the case of HTML). Moreover, occasionally files are missing or fail to download when a link is provided.

In addition to the idea that the mere fact of data's availability is enough, outside of any consid-

⁴⁸<http://www.mahaagri.gov.in/>

erations of standards, the researcher several times encountered the idea—particularly in the context of the two state government departments which this report addresses—that some data, however seemingly inoffensive, could or should not be trusted to the “common man,” and must remain safeguarded by the state. In one instance, for example, Thorat suggests that data concerning crop pest infestations cannot be made immediately public because it would cause mass panic among farmers, and it is the state's responsibility to ensure stability among its constituents. Alternatively, data on weather catastrophes, he says, might cause serious market fluctuations if released immediately and without due caution. That these kinds of data might benefit farmers is glossed over in favor seemingly of maintaining a certain level of control.

Tamil Nadu Agriculture Information Service Network (AGRISNET)

While on the surface the Maharashtra and Tamil Nadu's Agriculture portals look similar, examining the differences can show how the a variety of techniques in creating an e-governance project will affect opening the data later on.

As stated before The Tenth Five-Year Plan had several e-governance schemes in it this included the Department of Agriculture and Cooperation to create the Agriculture Resources Information System Network (AGRISNET)⁴⁹ scheme. This required each state to create an agriculture knowledge portal for farmers so they can get information electronically. Tamil Nadu's AGRISNET has received some attention for its innovative way of distributing agriculture information.⁵⁰ Like Maharashtra's system there is an abundant wealth of information on the site for farmers that they can also get through SMS services, district offices and CSCs.

Run by Tamil Nadu's Agriculture Department the portal has received 368589 hits in total (there is a live counter on the bottom of the website) and is a system set up to keep all district offices inputting data systematically and in a standardized fashion. The data created in the field is put on the website where it can be viewed in real time. You can drill down to the village level and find soil quality, where to buy seeds, weather information, and subsidy information for every registered farmer.

“Transparency is maintained because you are putting all beneficiaries online. Farmers buy the supplies from a depot then immediately it is online updated in real time. Exportable into excel options so people can print it out.” Explained Thiru P. Venkatachalapathy Agricultural Officer (GOI Schemes)-I.

The objective of the site is to promote having farmers populating data by working with field workers using CSCs and Agriculture field offices. The data created by this system is directly from the audience who uses it. So that the data accurately tells the department what is happening in the field.

They also took the time to bridge the gap of terminology that could cause confusion between government and farmers. For instance AGRISNET looks at soil quality as Nitrogen Phosphorus and Potassium (NPK levels) but farmers don't use this term so when soil quality is being determined the print out the farmer gets not only tells them the various NPK levels but also the types of fertilizer they need to buy to maintain quality soil levels.⁵¹

⁴⁹ <http://www.tnagrisnet.tn.gov.in/website/index.php>

⁵⁰ (<http://indiagovernance.gov.in/bestpractices.php?id=711>)

⁵¹ Soil Fertility Calculator (<http://www.tnagrisnet.tn.gov.in/website/Downloads/FertiliserCalc.xls>)

Also there are several SMS services as well so that farmers can get information on a variety of services and more services are being created. Also scheme information is available such as effectiveness and implantation information, however, this information is for internal use only. The site is currently used a way for the Agriculture Department to be better informed and to improve communication with the field.

They are hesitant when asked about making the data available. There is sensitive information that would have to be scrubbed and also this data shouldn't fall into the wrong hands. At the end of the day it is also not the Agriculture Department's call to make this data more open. Even though you can get reports printed from the site if you want the meta data or bulk data to do analysis you have to ask for permission. This data is too precious to share and they can't just give to everyone. You can see it in many cases but can't use it to remix or rehash. Some information is downloadable such as Seed Availability which you can download into .xls but not in bulk or in other formats.

When asked whether they hope to have open data from sites such as AGRISNET Tamil Nadu's IT Secretary Santosh Babu, replied there are a lot of concerns and privacy issues. "I think more openness is necessary but everything is online so it is free for people to see." The consensus is that that should be enough.

But it is clear that real time updates on a dynamic website are better than PDFs. If more e-governance initiatives create sites with real time updates and intern infrastructure that can support those technical solutions it would make moving to open data easier.

Kerala IT Mission

The Kerala State Information Technology Mission (KSITM) ⁵² is located in Thiruvananthapuram the capitol of Kerala and according to its website "is an autonomous nodal IT implementation agency for Department of Information Technology, Government of Kerala which provides managerial support to various initiatives of the Department."

Kerala's IT Mission primary responsibilities are

- Interfacing between the Government and the industry
- Interacting with potential investors
- Strengthening the IT / ITES industry base
- Holding promotional campaigns for hard selling the state.
- ICT dissemination to bridge the digital divide
- E- Governance
- Developing Human Resources for IT and ITES
- Advising the Government on policy matters.

In order to accomplish the above objectives they went about standardizing data and creating a central depository for the information. "The first step for e-governance is to have it all in one place without that we couldn't create our tools" said Mr. Sabarish K a Mission Coordinator. They also forged partnerships with other organizations in order to bring data to people directly. They worked with the Indian National Centre for Ocean Information Systems (Incois) in Hydrabad that does satellite ocean monitoring services to put this information up on digital boards in fishing villages so people can get weather and pertinent fish information.⁵³ They converted their illiteracy

⁵² (<http://www.itmission.kerala.gov.in/about-us.html>)

⁵³ <http://www.incois.gov.in/Incois/incois1024/index/index.jsp?res=1024>

kiosks and infrastructure into ICT educating kiosks to teach people how to use technology. Kerala has implemented 41 mobile services e-governance schemes that text you information as diverse as what hospital to go to if you are having a heart attack to how to track a hit and run driver.

“People who have data will demand more because they know how helpful it can be. They ask for more data and more accurate information.” When asked if they will make this data open Sabarish said “we aren’t there yet. We are still in 1.0 but one day we can move and get there. Attitudes have to change and that will come.”

Kerala’s literacy rate is the highest in the nation and makes a huge difference in teaching people how to use e-governance. Sabarish said that it doesn’t matter because it is all about how you present and package information. You still have to educate people and literacy is important but you need to package information in a way that makes it usable and effective with people.

The state level is not the only one moving to make information more available. Thiruvananthapuram Corporation has also made moves to be transparent. The accounting office puts the city’s budget up every month.⁵⁴ Kerala has decentralized some power down to its cities and panchayats. Thiruvananthapuram has started putting up its accounting information and while these aren’t open formats the information is updated every month. This is to fulfil the JNNURM requirements to be more transparent.⁵⁵ Even though the files are in .pdf they contain all whole budgetary picture for the city. There is new software being implemented to update the site and add more information faster. While the Secretary feels that decentralization of power will empower the city to collect more information about how the city runs and work to improve that he also feels people need to become more active as stakeholders in the city. And while the budget information is online for people to see few people look at it and use it to interact with government.

Kerala presents an interesting picture of organized e-governance and transparency initiatives that have little focus on open data but has dedicated itself to organize its data in a standardized way to make way to make e-governance a priority.

National Spatial Data Infrastructure (NSDI)

The National Spatial Data Infrastructure (NSDI)⁵⁶ created the India Geoportal to move toward the idea that data needs to be linked in a network and have an infrastructure so that it is most useful. The hope is to have data available “to meet the needs of users, planners, policy makers, industry and academia and to reap the benefits of higher-end technologies like GIS, GPS, high resolution satellite sensors, a single source of information infrastructure is a long felt need. Moreover, fast computers and ICT technologies provide ways and means for supplying spatial data to the users on their desktops.” So NSDI created the Geoportal to collect Meta data, related to special information, and catalogue the data sets on a portal. The portal now has made partnerships with 19 agencies who list their Meta data on the site and which people can find and then contact the agency to get. The site has had over 25,000 hits so far. There is no data actually hosted on the site you have to register to get it and pay money to the agency.

The program is run by Dr. R Siva Kumar and he is very dedicated to the idea of open data and would like data to be more available but he says there are many issues before data can be free. “We have set up a data sharing concept we have provided them with a particular standard. Bringing people together is an important requirement”.

⁵⁴ <http://www.corporationoftrivandrum.in/ml/pages/general.php?p=276>

⁵⁵ <http://www.corporationoftrivandrum.in/jnnurm/>

⁵⁶ <http://www.nsdindia.gov.in/nsdi/nsdiportal/index.jsp>

The catalogue is a series of drop down menus where you can pick what agency and then what Meta data you want. It then tells you how to get it and how much it would be. You have to contact the agency to get the data and must be a registered user. The data is also not free and has to be purchased from said agency. This is far from open but the first time the central government is collecting a list of data that exists.

Right now there aren't a lot of users and there still needs to be more data available. The data is still not open but the hope is that one day it will be by working with partners and coming up with a plan that can give them the financial needs to improve data but make it manageable. They are relying on the department to uphold accuracy.

One of the debates had been whether data should be prioritized by what potential applications that government defines or be released with no real guide. What would be the most useful for people? The answer is some mix of both but in India in particular it seems that the government's instinct to define applications for people to use is correct. And the NSDI hope to have data available and hopefully if the NDSAP is passed open data available then other players and come in and define applications as well.

Organizations That Work With Government Case Studies

This section is about groups that are either private-public partnerships, exist because of NKC recommendations or are working with civil servants to create a more open atmosphere. This is grouped because it is important to see how groups are working with government actors to change attitudes by either dealing with them directly or creating tools with open data that have proved their usefulness.

India Water Portal

The India Water Portal, based in Bangalore, defines itself as "an open, inclusive, web-based platform for sharing water management knowledge amongst practitioners and the general public. It aims to draw on the rich experience of water-sector experts, package their knowledge and add value to it through technology and then disseminate it to a larger audience through the Internet."⁵⁷ The portal gets 3000 visits a day primarily by water experts and stakeholders but the goals is to expand reach eventually.

The IWP embodies the best practices evident in the portals established on the recommendation of the NKC. Active since 2007, the IWP, which runs on the open source Drupal software, actually consists of a series of portals providing information in English, Hindi, Kannada—the local language of Karnataka, the southern Indian state where the portal is based—as well as on the subjects of water conflicts and sanitation. The IWP also maintains a separate portal for schools providing educational materials on water issues for students and teachers. The flagship portal provides data and statistics on rainwater harvesting, agricultural issues related to water, drinking water, water bodies, and urban water.

Deepak Menon, who coordinates the Hindi, Schools, Sanitation and Conflicts Portals within IWP, says that the idea behind the portals is that they should each have an identity of their own. Each language portal, for example, features content not found in the others, based on what information is available in what language, converting what might be seen as a weakness into a localizing (and flavor-enhancing) strength. The IWP does not seek to translate materials, but does encourage other individuals and groups to translate the information that it stores. Probably its greatest

⁵⁷<http://www.indiawaterportal.org/about>

weakness is that it is accessible only as a web-based technology, and thus only by a certain class of people. But Menon does not see this as a major impediment to its mission, because its information is oriented more towards local NGOs who can analyze and interpret the data and information which it provides, as well as students and teachers learning about water issues.

Although the IWP was setup at the instigation of the national government, which Menon says has been quite supportive of its work; he says that the portal has not had a tremendous amount of success getting information from government ministries and departments, which are the biggest producers of knowledge on water and sanitation. Exceptions include the India Meteorological Department's (IMD) datasets from 2004-2008, which the portal acquired from an NGO that filed an RTI request for the data and then provided it to the IWP; station-wise daily rainfall data for all districts of Rajasthan from 1973-2008, which that state, known for its robust public data practices, freely discloses publicly; and the archives of the National Institute of Hydrology. The majority of the IWP's content, however, comes from NGOs and a small subset of users who upload data themselves. (The portal is a free platform, and anyone can add content on their own.) Most datasets are in electronic, machine-readable form, and are viewable in web browsers as well as freely downloadable.

Menon does say, however, that the IWP has not found the government blind to the idea of sharing information, and what government information that the IWP does host has been freely offered, with the exception of the IMD datasets. Moreover, he says, the IWP would like to work far more closely with the government to correlate data and put it in more usable formats, such as machine-readable spreadsheets as opposed to the scanned PDFs via which the government often provides data regardless of its nature. And he is optimistic that such collaborations will indeed happen.

The culture around government data is changing, he says, and "you find [government officials] talking about data in a far more open way than in thinking about it in terms of an internal...report." Indeed, he believes that "it won't be that difficult in the future to get something that you need" without resorting to avenues like the RTI Act.

In fact, the IWP has never itself filed an RTI request, although Menon says that increased use of the RTI across India may be a factor in the shifting government practices around data sharing. In his experience, though, getting data has been a result of institutionally "having good relationships with people [in government]," which he says may be more challenging for some organizations than others. He regards the IWP's generally warm relationship with the government, he wishes that the government was more forthright with its information, being proactive rather than reactive. Data has to be out there in order to see what people can do with it. Give us the data and then we can deal with everything else. It is important to view data in terms of who the users are and what the usage could be. Eventual users of data can go beyond what government imagines.

One of the biggest problems that the IWP faces in getting information from the government, according to Menon—and it is a problem often faced by people searching for government data, including in the course of this report—is that almost all government data is managed by the NIC, which is responsible for designing and maintaining many government websites and which often employs lax standards for navigability and searchability. And often, he says, it is not clear who one should contact to report problems with websites.

Regardless of existing problems with government disclosure and the websites through which it provides information, Menon does not believe that a national open data portal is far off in the future; he thinks that it will happen, but contingent upon when and how strongly society demands it.

YASHADA

Yashwantrao Chavan Academy of Development Administration (YASHADA) is the Administrative Training Institute of the Government of Maharashtra, and works primarily to train government officials, civil society actors, and other stakeholders in areas like good governance, sustainable development, and information technology. It also works with various stakeholders to sustain projects at the village, block, and district levels by facilitating engagements with local politicians and government officials. Currently, it is involved in establishing data collection formats at the village level for purposes of microplanning, which it is collaborating with UNICEF to promote. YASHADA has trained 60,000 officers. They do not go into villages for RTI but the officers they train do and so their reach can be seen in nearly every village in Maharashtra.

YASHADA, which generates its own revenue, was named by the Department of Personnel and Training as an RTI resource centre, as part of which—as with all of its training projects—it has worked with an alliance of trainers to increase awareness of RTI in villages throughout the state. They plan to reach 40,000 villages during RTI week which is the 5th to the 12th of October.

As director of the Center for Public Policy at YASHADA, Prahlad Kachare, who is visiting at the academy from the state revenue department, has worked for the past year to promote use of the RTI in villages throughout Maharashtra. Kachare is unique in having been on both sides of the RTI, as part of a department required to disclose information and as part of an organization tasked with enabling people to request that information be made available to them. In both cases, he has noticed that the government's information practices are lacking.

While departments are compliance oriented and will put PDFs on their websites, the data presented is neither truly accessible nor updated regularly, and moreover there is reluctance and a lack of motivation on behalf of bureaucrats to disclose any more than necessary. Partly, he says, this results from the RTI—there may actually be less proactive disclosure “because there are more reactive inquiries from citizens ... [who now] know that they have some right [to inquire].”

Open government data, Kachare says, is possible in India and even necessary, because it will reduce the amount of requests with which government departments must deal and require the government to submit to greater scrutiny. That aside, the key problem in propagating such a policy from the perspective of YASHADA is not necessarily that officials do not have the capacity to deal with disclosing information, but rather that officials lack proper training, and that translates into a lack of institutional awareness about collecting, arranging, and disseminating data. Nor are the grassroots workers who might ultimately make much use of open data coordinated in such a way that they could really benefit from it. The issues—whether facing government or civil society—are primarily about resources, mindsets, and organizational behavior, as well as inculcating a sense of accountability in officials.

In addition to training, Kachare suggests that mindsets and organizational behavior might be in-

fluenced not only by training and capacity building exercises but through an incentive-based system protecting and rewarding officials engaging in best practices. Also important is a focus on holding top-level management accountable for the data practices of their subordinates.

Although one might expect that an official tasked with training people to request information from the government would believe in the power of information, Kachare does not believe that the government loses such power when it opens its data. Those bureaucrats who fear losing their power, he says, are mistaken. “[They think] they will lose their power, their authority. In fact they don't. The more information you give, the more respect people give you...The more you inform people, the less people will ask you, the less people will doubt you. The system [runs smoothly and is of utility] to people.”

Regarding ensuring the quality of information—another task of organizational management—the issue is moot until the government actually starts providing more information.” Let the information come out, and then we can talk about quality,” he says.

Aside from problems related to the government, Kachare says that the mindset of the people also must change for an open data policy to be effective. “At the moment, people do not believe in [electronic information]. They believe in signatures,” he says. To change the people's mindset, organizations must approach them directly, not just via their representatives at the panchayat or other levels. Part of that challenge also is “how to move people from individualistic requests to [larger issues].” People do not understand and cannot interpret data with a view to participate in governance. I feel open data needs to be searchable/retrievable/accessible database. And then we can think of training citizens including villagers how to search data of their use and how to use it. Making data available in searchable mode should be governmental responsibility and once it is available CSO's can take over task of training people to use it in a responsible manner for participative governance.....or say good governance.

He suggests that to effect a truly useful open data policy, organizations like YASHADA will have to follow the same model that they have in promoting the RTI, which involves traveling to villages and informing people via various forms of media—and in their local language—about the information to which they have access and the ways in which they can use it. Few villages now, he maintains, are aware of all the resources contained even just within their own boundaries. Open data may help change that if people are actually equipped to use it.

Reserve Bank of India

In 2002, the Reserve Bank of India (RBI) began using an internal database consolidating data on various aspects of the Indian economy as well as select international economic data. In 2004, the RBI decided to open its database to the public. Since then, more than 5 million users — at this point, about 20,000 every month — have accessed the Database on Indian Economy: RBI's Data Warehouse (DBIE).

The DBIE is arranged by subject areas including macro economic aggregate data, financial market analyses, macro monetary management, currency management, RBI financial statements, and banking sector information, among others. The website by which users can access the DBIE allows users to view pre-formatted reports in these subject areas by time period, and to download reports in CSV, XLS, or PDF format. (The RBI is currently working to implement XML downloads.)

The website also allows users to create reports based on selected variables and time periods chosen via both simple and advanced query systems, and to view metadata for all variables included in the DBIE.

Ashish Jaiswal, assistant adviser in the RBI's Operational Analysis Division, says that the RBI has focused on making its data as accessible as possible. "When you're putting such a large volume of data out, you have to arrange it in a manner that is easily accessible by the users," he says, which is why the RBI organized its database by subject. Part of accessibility, he says, is making data available online as soon as it is ready for public consumption. Once it made the decision to open its data, the RBI has endeavored to get its data online as quickly as possible, taking a staggered approach to introducing material in order to provide the most data possible at any given time. The bank has also focused on maintaining its data, updating the DBIE every day, and it encourages user feedback, to which it responds on a weekly basis. All these things, suggests Jaiswal, provide a model for how government departments might organize their own open data schemes.

According to Jaiswal, the DBIE "serves the need of transparency and gives confidence to investors to invest and [helps them decide] on what to invest in India." Economic data is an area where India scores higher than its developing rival China, he says, where investors may not find much data, and he draws a link between the RBI's transparency and India's overall global economic competitiveness. "We are ahead in this field where we share a lot of information with the public...We are very much transparent. We don't hide the information ... [and] we are competing in that way."

In addition to providing information to investors and benefiting the Indian economy, Jaiswal says that there is a "big battalion of researchers" both within and outside of India to whom the RBI is catering with the DBIE, including students and academics who otherwise would lack access to data that previously has been published only on paper and accessible in few locations.

Response to the DBIE has been positive, not only in India but from other developing countries, which have sought to model their own economic databases on the DBIE. "We are the pioneers. So in fact other banks have actually come to us — a lot of African banks, a lot of banks from South-east Asia," Jaiswal says. "In fact the Bank of Thailand have built their database on our model."

Unfortunately a lot of the DBIE's functionality, including advanced visualizations and other means of analysis, is available only internally due to bandwidth constraints. The RBI would like to make this functionality available in the future but is unsure when it will have the technical capacity to do so.

National Institute for Smart Governance

The National Institute for Smart Governance (NISG) is an NGO incorporated as a public-private partnership in 2002 with the assistance of the National Association of Software and Services Companies (NASSCOM), the Indian software industry's public policy consortium; the Government of Andhra Pradesh, a state in southern India with a long history of e-governance initiatives; and IL&FS, an Indian company focusing on infrastructure development and finance. NISG, which provides consulting support to both central and state Governments in India in developing and implementing e-government projects, is chaired by Mr. R Chandrasekhar, secretary of the Department of Information Technology. Much of NISG's work relates to implementing India's National E-Governance Plan, which seeks *"to create the right governance and institutional mechanisms, set up the core infrastructure and policies and implement a number of Mis-*

sion Mode Projects at the center, state and integrated service levels to create a citizen-centric and business-centric environment for governance.”⁵⁸

Recently, NISG has been working with the Unique Identification Authority of India (UIDAI) to develop a conceptual framework for the country's proposed Unique Identification card. Considering its role in working on the UID project, as well as the link between the UID and open government data envisioned by Mr. Pitroda, it is not unlikely that the NISG might also provide high-level consultancy services related to open government data planning and implementation.

T. Vijay Saradhi, vice president of NISG, suggests that open government data in India may need to serve a different function than it does in other countries like the United States and the United Kingdom, where he says open data provides citizens with a feeling that the government is open. “I don't think that [feeling of openness] is so important in the Indian context,” he says, where “what is more important is how is it going to be useful to people—and whatever schemes and programs you have, how are they percolating down to the actual beneficiary. So openness in the government should be focusing on specific items that the government is trying to do for the people,” rather than on fostering a general sense of openness or transparency.

As an example, Saradhi points to the central government's flagship schemes like the Mahatma Gandhi National Rural Employment Guarantee Scheme (MGNREGS), whose implementing ministries and departments must report to a delivery monitoring unit within the Office of the Prime Minister regarding the outcomes of their services. The important information in these reports, he says, is not just about how much money the government is spending and where, but rather about what outcomes are being generated. It is not transparency for transparency's sake that is important, but transparency with an eye towards accountability.

Satyajit Suri, general manager at NISG, elaborates on the point that the US and UK models may necessarily have different objectives than an open data policy in the India. There are three main objectives of Data.gov, he says: participation, transparency, and collaboration. India might have the same notion of transparency, but as for participation, he says, “There the whole idea is that the government throws up some data, and citizens can act on the data and do something with it. Are we [in India] in a participative or collaborative mode? I'm not sure. We have to think about what that means in our context.”

Moreover, Suri asserts that “The US and other western countries are very public policy-focused countries. They tend to use a lot of data—[there] job data can change stock markets overnight. Let's face it, it doesn't happen here. But data is very sacred for some reason over there, and they tend to believe that they can do a lot with data—especially those in public policy. And in India I'm not so sure if our policy makers rely so much on data as much as they rely on more social information.”

Indeed, Saradhi stresses that “[w]e are looking at it always from the citizen's point of view. So we look at it and say what is the benefit at the end of it—is the transparency there to account for whatever has been spent or proposed by the government?” He cautions, too, that “[India's] level of maturity to use this information for public policy [is lacking] ... If you go to local government, it's very difficult to find someone who can use data. Which is not a very proud thing to say, but that's a fact of life. so if you give 'Data.gov' to policymakers in India, I'm not sure that it can be [used] for the purpose it's intended [in the US].” Furthermore, he says—echoing a refrain often heard in the course of researching this report—open data might be detrimental to the Indian political system in ways of which no one is yet fully aware.

Suri, at least, does believe that in practice the US and UK can serve as models for India in opening its data, in part due to the simplicity of their initial approach. “[United States Chief Technology Officer] Vivek Kundra did something great with the US,” he says. “He just took a lot of these reports which were unusable in some sense ... and he made them accessible in a way where somebody can mash it up and create a lot of analysis on top of it. That's the beauty of data.gov.”

India, Suri says, can likewise place its existing electronic data online. And while he does not think that India need open everything, he likes the idea that India might encourage innovation upon data as with the US and UK models. “We need to open something that can really show [people that] this is useful for [them] as [citizens] of the country,”

⁵⁸http://india.gov.in/govt/national_egov_plan.php

he says. [And] if I can create a lot of analysis on top of it, that's great."

Innovation aside, however, Suri suggests that the promise of open government data in India is not so much transparency or accountability per se as it is the possibility of empowering citizens at the lower political levels of society—that is, in their villages—to take greater control of their lives based on the provision of public data on which communal choices might be made in an informed and deliberative manner. Another standard refrain heard in the course of this research is that in India there is no one standing up for the common man, that the government pays him no heed in its decision-making process and denies him the right to self-determination. And data, many say, including high level central government officials—at least data that sheds greater light on the relationship between the citizen and the state—may limit to some extent the power of the centre over the periphery, where more than half of India lives in a more or less undeveloped world. There is great faith in India, especially in light of the RTI Act, in the power of knowledge and of the ability of rural Indians to take control of their lives once they possess certain knowledge about what is owed to them—and moreover about what possibilities may be hidden in their surroundings. The problem that this highlights is that in such a large, heavily bureaucratized state as India, where decision-making has been centralized and the bureaucracy is a world unto itself, knowledge is a key to authority that many bureaucrats (and the issue seems to be with mid-level bureaucrats, not high level officials) do not want to cede.

Civil Society Case Studies

This section deals with the way data affects the conversation with government and the ability to promote accountability and transparency. How organizations are using government data and how making it open can help their work.

IndiaGoverns Research Institute

IndiaGoverns Research Institute⁵⁹ is a relatively young non-profit organization (currently funded by a fellowship grant from Echoing Green) that works with development data, and seeks to make that data more useful for policymakers, researchers, civil society organizations, presspersons, and ordinary citizens. To do this, they have chosen to focus on a single state (Karnataka), and have extracted developmental data from a variety of government departments using various means (by filing RTI requests, by contacting officials by e-mail, etc.). The developmental data they seek to gather are of the following six categories: education, health, water, roads, agriculture, transport, and a miscellaneous "general" (which includes information on population, schedule castes & tribes, electrification, forest area, etc.)

The organization, headed by Veena Ramanna (and currently run largely on volunteer energy), envisions itself as a change agent, helping shift political discourse from one of rhetoric to one of substantive issues. According to their website, "IndiaGoverns Research Institute aims to make development data matter in two ways: make the development discourse between elected representatives and citizens better informed and backed by specific data, rather than only relying on perceptions; and enable citizen groups and elected representatives to use such government data to strengthen their demands for greater government intervention on development issues relevant to them." It aims to do this by foregrounding developmental indicators and statistics down to the taluk level, along the lines of political (electoral) constituencies rather than administrative divisions.

While elected officials such as Members of Parliament, Members of Legislative Assembly, and Gram Panchayat members are directly accountable to their electorates. However, the bulk of the data that is gathered by the government is gathered along administrative boundaries, rather than electoral ones. Thus while this allows for comparisons to be drawn between different administra-

⁵⁹ <http://beta.indiagoverns.org/>

tive divisions, the data needs to be reorganized if comparisons are to be carried out along lines of direct accountability. This is made even more difficult by the fact that different departments might collect the same data (information on schools might be collected both by the Revenue Department as well as the Department of Education), but they might be collected along different boundaries. This point was reinforced by Gautam John of Akshara Foundation who spoke about the disparity of information geographies as one of their largest problems, since educational information boundaries don't match health information boundaries, which don't match parliamentary boundaries. This (re-)mapping is being done by these groups themselves.

According to Ms. Ramanna, often this information is not only difficult for ordinary citizens to access, but is equally difficult for politicians to get hold of, as well. This accounts, she notes, at least in part, for political discourse (seen starkly in the form of election speeches — a number of which IndiaGoverns have recorded) on the basis of rhetoric (even if one of development) rather than on facts and indicators of development.

IndiaGoverns hope to present this reorganized and redesigned data (with accompanying graphics) with helpful comparisons (interactively between individual constituencies, as well as with averages) to Members of Parliament, Members of the Legislative Assembly, the press, NGOs, and others. They see themselves as a non-partisan body that will not get involved in lobbying, but merely as a facilitator of political dialogue through data gathering, data analysis and research dissemination. "There is a lot of interest and it's an important tool. They started to ask please do this for our constituency. They want this work," comments Ms. Ramanna.

IndiaGoverns is still relatively new and have just recently finished reports based on the data they collected.⁶⁰ So it is too soon to measure impact and the traffic to the site is minimal. However they plan to specifically get these reports to journalists and politicians so that they can use them to evaluate programs in time for the next elections. Thus, while it is not possible to consider the impact they have had, it would be interesting to speculate on the impact they might have. It is important to note that while all the information that IndiaGoverns makes available is directly sourced from the government (they do no data collection of their own, and don't rely on data from other sources), it is not possible to inspect the range of data that they are making available in any single governmental website. They are, in a sense, opening up the data on behalf of the Karnataka government.⁶¹

There is no master list of information that government collects. What seems to be happening is that inputs to programs are collected but there is a lack of follow up. Start measuring and rethink what you need to provide people with. "One of the things is if you don't know what data the government is collecting we can't make suggestions on what data should be collected" added Ms. Ramanna.

Because they are not collecting the data themselves, they make no claims about its accuracy. Thus, this may end up, exposing incorrect data to the public eye putting pressure on government for more accurate and better and robust data.⁶² And if a mechanism is built into IndiaGoverns for

⁶⁰ IndiaGoverns report generator (<http://beta.indiagoverns.org/summaries/>)

⁶¹They have not studied the copyright implications of reusing government data, but as has been examined in a different part of this report, that might not be as much of a concern.

⁶²While there are other instances of incorrect data being recognized because of RTI requests, it is not often that

identification of each data source and communication with that department, then that could slowly end up in feedback to the departments and proving Linus's Law that many eyes make all bugs shallow. The organization is also willing to provide the raw data that it has collected for bulk downloading or through open APIs. This work will change the perception on data by elected officials which could in the end have an effect on collection and distribution later on.

Centre for Budget and Governance Accountability

The Centre for Budget and Governance Accountability (CBGA) is an NGO located in New Delhi that advocates for "transparent, accountable and participatory governance, and a people-centred perspective in preparation and implementation of budgets."⁶³ The centre was founded in 2002 as an initiative of the NCAS in Pune and is now an independent organization analyzing national and state budgets "from the perspective of the poor and the marginalised."⁶⁴

All of CBGA's work falls into one of three categories: budget advocacy, where the centre seeks to promote equitable public policies; research and analysis, where it produces major research studies and manuals and primers on major policy issues, as well as a tri-annual newsletter in both English and Hindi simplifying policy issues for general consumption; and capacity building, where it has partnered with other NGOs and UN agencies to facilitate decentralized budget analysis and empower the grassroots through demystifying government processes. The information they collect is primarily dispersed through studies, reports, publications and the website. Each publication has 1500 printed copies. Priority list includes courier to policymakers, targeted MPs, media people, academics, civic society orgs. The list is about 1000 to 2000 people long.

Subrat Das, executive director of CBGA, identifies four key problems that his organization encounters in working with government budget data.

First, data sources on public expenditures and revenues are scattered. If someone wants to figure out how much India spends each year on health, for example, she must consider both national and state sources, as well as various institutional sources at each level. "There are multiple institutions dealing with these processes, and every institution maintains its data as a silo," Das says. "And there is not much effort to have coordination ... and present the data in an organized manner."

Second, there is a serious time lag in most budget data in India. For example, actual budget data (as opposed to estimates) for fiscal year 2011 is only available at the start of fiscal year 2013. "It is possible for the government to make the actual expenditure and revenue figures available ...well in advance," Das says. The central government, he says, has begun to make current data available through the office of the Comptroller and Auditor General (CAG), but state governments have not yet made serious efforts to present their data in a timely fashion.

the replies to RTI requests are this well organized and presented. For example, Shailesh Gandhi, when he was still an RTI activist, had asked for information pertaining to murder rates in Mumbai from 1983 to 2007. He found that the data could not be accurately correlated with the increase in Mumbai's population in those 25 years. However, that finding never got publicised properly, and was not backed up by the kind of comparative analysis that organizations like IndiaGoverns or Akshara Foundation are currently doing.

⁶³"Overview," Centre for Budget and Governance Accountability. http://www.cbgaindia.org/about_overview.php

⁶⁴Ibid.

Third, it is often difficult or impossible to compare datasets—even when they are ostensibly reporting the same variables—due to varying and unarticulated methodologies. For example, there are three primary sources of data on government expenditures in India, including the Indian Public Finance Statistics⁶⁵ and the Economic Survey⁶⁶, both produced by the Ministry of Finance, and the Combined Finance and Revenue Accounts of Union and State Governments⁶⁷, produced by the CAG. Figures across these sources are not always comparable, as where the Economic Survey reports that spending on health is 1.4 per cent of GDP, but the Indian Public Finance Statistics report that spending on health is only 1 per cent of GDP. The reason, Das says, is that perhaps the Economic Survey includes spending on water and sanitation as codeterminants of health, whereas the Indian Public Finance Statistics list them separately. But the Economic Survey does not clarify whether this is the case, leaving researchers guessing as to the actual amount of health spending.

Finally, not all budget data is reliable, especially where it concerns expenditures. In order to facilitate implementation of national flagship schemes like the NREGS and NRHM, central government ministries have often bypassed state budgets and sent funding for the schemes directly to the autonomous bank accounts of their implementing agencies. The CAG is not mandated to audit these funds, and since the funds are not handled by state treasuries, they are neither subject to state auditing procedures. In many cases, then, the central government actually does not know how much money has actually been spent on the ground as opposed to disbursed.

There are several positive outcomes that the CBGA has observed from their work. CSO, International/national NGOs are more responsive and sensitive to budget issues, and will include budget analysis in advocacy and work. More people are paying more attention to the discourse around the budget. Especially the budgeting for minorities, dalits, women and children there has been an influence on strategies for union government on budgeting issues, better outcomes and transparency. Mr. Das went to say that there has been more transparency in budget documents and the union government is making more information available.

Citizen Matters

The media as stated before can benefit greatly from open data. There is a great advantage in grounding journalism in data that can be used to sort through the rhetoric that politicians use. Open data will allow more people who are not part of the “main stream media” to write stories and ask the relevant questions without having to file expensive RTIs or rely solely on the relationships built with bureaucrats.

Citizen Matters “a Bangalore focused, citizen-oriented newsmagazine, covering city public affairs, community and culture” is published by Oorvani Media, an independent local media firm founded by Subramaniam Vincent and Meera K.⁶⁸ According to their website their credo is to “encourage citizens, professionals and journalists alike getting involved in city affairs, reporting, collaborating on, analysing, and proposing ideas and solutions.”

Mr. Vincent has an engineering background and worked both in India and in the US. He switched to communications and journalism when he co-founded the national magazine India To-

⁶⁵Indian Public Finance Statistics. <http://finmin.nic.in/reports/ipfstat.html>

⁶⁶Economic Survey 2009–2010. <http://indiabudget.nic.in/es2009-10/esmain.htm>

⁶⁷Combined Finance and Revenue Accounts of Union and State Governments. http://www.cag.gov.in/CFRA/CFRA_year.htm

⁶⁸<http://bangalore.citizenmatters.in/>

gether, which covers many national issues. ⁶⁹ “We live in a data starved society,” commented Mr. Vincent, “we are still not a data looking society but more of an opinion society. Data is part of follow up question not a starting point for news.”

When asked what kinds of data the City of Bangalore collects. Mr. Vincent explained that there exists no structure internally that allows the city to collect data in a real way. Also the cities don't have enough power to see data collection as useful to begin with. If there is information it is around projects such grid separators – flyovers and underpass – that are structures used to decongest traffic. Citizen Matters wrote a story about these construction projects and wanted to get a list of the upcoming grid separator projects. To get a list it took a month. When they did get a list it was through an official at the BBMP but it wasn't a formal document just a list on a word document. The lack of proactive disclosure regarding projects like these is that there is a lot of damage and waste that can happen as these structures are built and the city doesn't publish information until the entire project is finished which is too late to make any pre-emptive recommendations.

Tendered projects are put up on the Bangalore city government's website, and these are post-approval. However, project feasibility studies usually contain information that ought to be made public before approvals are given. Such studies (for e.g. about the grade separator projects) are rarely released online.

Even RTI requests can be difficult. If you don't ask the correct question with the exact terminology the government uses they will not fulfil the request. In response to this there are people who have mastered the art of filing for RTIs to get the optimal response. In the early stages, Citizen Matters had their applications vetted by these experts before they go to the city to make sure everything is correct so they minimize the chance of a run around later.

Issues around quality of data are rooted in poor bookkeeping. It is not uncommon for there to be multiple documents for the same list of projects. There could be three different documents from different times instead of updating one document continuously. There are many incompetency's in data management. “When you access things that come out of government offices, you find that they don't have a single coherent way of communicating about their projects. No need to. No one is sitting with them saying do this. No way for this order to actually happen,” replies Mr. Vincent.

“When you speak to officials they don't use data to talk about a program or situation.,” added Reporter Navya P.K. “Very rarely do I see statistics being given out by them.” Even when the city provides information it is so incomplete it is not useful. Also officials tend to be more open with data that is from studies from other organization that were done with a large amount of funds, especially scientific data. However with data that shows something that officials are doing or were responsible for they are less willing to share. For example they have been doing a lot of tree cutting for the road widening but no one has a list of how many trees and where and there is no indication that they have collected that data.

In addition to classical narrative based journalism, Citizen Matters wants to root all major government and society related stories in data. “We are data people that's our background and we

⁶⁹ <http://www.indiatogether.org/>

would like more local data but often times there simply isn't any" laments Mr. Vincent. More data available will mean that there will be more information for journalists and the like to pour through which will give them a better base to engage politicians and officials.

Transparent Chennai

Transparent Chennai⁷⁰ launched in October 2010, by the Centre for Development Finance, to collect city data, map it, make it open for people to find, use and to work with citizens to enable them to create their own data. According to their website Transparent Chennai "aggregates, creates and disseminates data and research about important civic issues facing Chennai, including those issues facing the poor." They also "work closely with individuals and citizens groups to create data that can help them counter inaccurate or incomplete government data." They get around 100 hits a day but given how new they are it is too soon to tell what is their impact is as of yet.

Nithya V. Raman is the founder and the Project Lead and has a background in Urban Planning. The project is staffed by researchers and programmers from CDF, consultants and a regular supply of interns from local and international universities who help get data, verify it and then map it. They work to map any information the government provides gathered through RTIs and available public documents so that people can use to engage with government and create better policy around needs and vulnerable populations.

The site allows the user to layer maps on top of each other to get an interesting idea of where the overlaps are. The site also enables users to input information on select layers, and invites participation from citizens to create new layers and collect data for new and existing layers. Information from in-depth research on particular civic issues available on the site under sanitation, slums, road safety, solid waste management, electoral accountability, and inclusive planning. These categories are to show that there is a whole ecosystem to city planning and governance where data is needed to make better decisions.

Their work on mapping all the public toilets in Chennai is a prime example of the issues that they face with collecting city data and how better information can make an impact on policy decisions.⁷¹ First they asked for a master list for the public toilets from the Chennai Corporation. The city did not have a centralized list so they had to ask each zonal office for a list of public toilets in their area. After getting each list, it was determined there were 572 public toilets in Chennai. Afterwards they sent an RTI request to each zonal office for the list of public toilets, this list produced a total number of 715 toilets, a discrepancy of 143 toilets. "This is typical of data collected at the city level" comments Ms Raman, "They don't have accurate and systematic collection systems so there are discrepancies." Afterward each location was visited by Transparent Chennai in one zone, Zone 4, to verify there was a toilet and to georeference the location and to conduct a survey on the quality of the toilet and brief interviews with the users, and then all this information was mapped. This process of data collection and verification is similar for all sets of data they have, although certain features of the city like flyovers and reservoirs can be mapped using satellite imagery and do not require physical verification. The government provides almost no geo-referenced data beyond the Census. Then they put the data up on their website usually in excel, .doc and pdf formats.

Out of this work it was largely determined that public toilets are not placed where there is need for them; areas with a lot of foot traffic like bus stops and where informal sector workers work who do not have access to private facilities such as market areas. Comparing toilets data with

⁷⁰ <http://www.transparentchennai.com/>

⁷¹ <http://www.transparentchennai.com/research/sanitation/in-depth-research/>

data from Tamil Nadu Slum clearance Board of undeveloped slums in the city also reveals that most toilets are not located near these slum areas, presumably the areas of greatest need for sanitation. By looking at the data it can be clearly determined that the city has been given money to build public toilets and have to use it but the city does not use available data to determine where the greatest need for these toilets is.

“The lack of data in India enables government to evade responsibility to provide basic services to residents, and sometimes allows for the use of force on them, such as slum and street venter evictions” explains Ms Raman. She explains that they digitalized the Slum Clearance Board’s map of undeveloped slums in Chennai, and compared with a map of evictions created by asking citizens’ groups to mark eviction sites. Comparing these maps revealed that the city evicts many slums that are not even recorded or recognized by the city.⁷²

Transparent Chennai hopes their work can be used to empower citizens to participate in urban governance – including participatory processes like the City Development Plan (CDP), City Sanitation Plan, and other plan-making, and to interact more effectively with the government on a broader level. “There won’t be enough impact to policy and planning if it (data) sat on the internet” adds Ms Raman. So Transparent Chennai started having meeting offline with stakeholders to make people aware of this tool and what it can do for the people these organizations serve. The group also started working closely with specific civic organizations like environmental organizations, RWA and slum organizations to enable them to collect and represent data on a neighbourhood level that could help them better negotiate with the government.

Digital Green

Digital Green “aims to raise the livelihoods of smallholder farmers across the developing world through the targeted production and dissemination of agricultural information via participatory video and mediated instruction through grassroots-level partnerships.” Founded by Rikin Gandhi Digital Green is operating in 5 states and is actively expanding to more. They have produced 1654 videos, held 35227 screenings and have involved 57568 farmers.⁷³

They do this by teaming up with local CSOs who have experience with the local farmers in the area. They local CSO or in some cases field staff meet with farmers identify techniques and create a video. The CSOs are also in charge of setting up viewing sessions and providing the equipment to do so. Then when a farmer is implementing a technique or if they have questions about it they can contact Digital Green’s experts in the field and partners who can help them out. Digital Green not only has a large amount of video content but also their implementation data is kept on their website. Where you can see how many farmers saw videos, where and what the adoption of techniques was.⁷⁴

Usually they ask government for packages that could be relevant which include appropriate practices, information on how to improve yields, etc. This information is organized as a one pager provided by the government. Digital Green will take the fact sheet and try to localize it and demonstrate best practices, further adaption, and then ask local farmer to try that technique. Then go back and see what the results are and if the information the government provided is relevant.

⁷² <http://www.transparentchennai.com/research/sanitation/who-is-responsible/>

⁷³ <http://www.digitalgreen.org/>

⁷⁴ http://www.digitalgreen.org/analytics/overview_module/?geog=state&id=10000000000001

“Many of these partner organizations have gaps in knowledge and go to agencies where they get some information in form of fact sheets of data and that will be taken and localized in the format of these videos. But there is exchange of data going on. Taking the video and the data we have and sharing with the government agencies this helps inform their research and education curriculum in a way to better their programs” explains Mr. Gandhi.

People are basically getting government information plus other localized information through these videos and through Digital Green’s feedback mechanisms can take part in shaping information that government has and presents to others. “There definitely have been some changes to the fact sheets they give and curriculum. And ofcourse there is interest in which techniques are working and which are most popular. Government has interest in those things.” Mr. Gandhi goes on to say the problem is that “the data they provide is just too generic and not localized enough for us to use it. So we have to make it relevant to the communities we serve.”

Civic Hacking Case Studies

This section looks at the world of data use in India and what the current civic hacker projects are and what issues they have.

IndianKanoon

IndianKanoon.org is a legal search engine (‘kanoon’ being the Hindustani word for law) created by Sushant Sinha, a computer programmer, when he was doing his Ph.D. from the University of Michigan. IndianKanoon indexes judgements and statutes, with automated relational hyperlinking between different documents (for instance, when one case refers a statute, or to another case). One area in which the government has been most effective in making data available has been in the areas of statutes and judgements. However the problems that are faced in accessing these, for both citizens using the government site directly as well as for automated crawlers of sites like IndianKanoon point out how bad the situation is even in cases where the government has done the job well, relative to other areas. IndiaKanoon has had around 2.5 million pageviews a month and a half million of which are unique. Their audience seems to be general practicing lawyers and academic researchers but there is no outreach to any particular groups.

The Indian Government portal JUDIS (judgement information system)- judis.nic.in makes available reported judgements of Supreme Court of India and several High Courts. Likewise, the portal IndiaCode.nic.in contains all Central Acts. IndianKanoon scrapes both of them (at one point of time, JUDIS stopped publishing feeds, which would have made the job easier). The contents of JUDIS and IndiaCode are correlated and integrated together in IndianKanoon, thus making the website very useful.

Apart from those sources, IndianKanoon integrates (and in some cases, is planning to integrate) reports of the Law Commission, open access law journals, and other online legal repositories. A site like IndianKanoon.org is necessary because the official interface for searching through judgements is extremely poor, breaks often, and is user-unfriendly.

Mr. Sinha, who now works with Yahoo India in Bangalore, has run into problems with court authorities (from the Allahabad High Court) trying to prevent spidering of the judgements by using CAPTCHAs, requiring image-based verification of humanness. It is unclear why they have done so, and no contact details are available for the technology or technology policy team on High Court’s website. Sushant doesn't consider the site a success yet. There is definitely more information out there and some courts seem to be more proactive but there is a long way to go. He hopes to update the site and make the articles searchable. Right now there are no search capabilities or it's not free. The Lok Sabha debates are not in a searchable for instance. Sushant scraps the site for Hindi and English he isn't use RTIs for the data. Right now an API is not a high priority.

OpenCivic.in

OpenCivic is an API under development by Akshay Surve, a Mumbai-based social entrepreneur, “to liberate civic participation related data in a machine readable and re-mixable form that will allow developers and visualizers to in-

teract with this critical data and build engaging applications over it in India.”⁷⁵ Surve intends to partner with websites and civic organizations to provide citizens with access to government data via the OpenCivic API, which is not an end-user application, but a way of fetching certain data. Currently, Surve is working with AskNeta⁷⁶—a website seeking to connect citizens with their elected representatives—and GovCheck,⁷⁷ a website consolidating and analyzing data on elected officials from various sources such as the Lok Sabha (lower house of Parliament), Rajya Sabha (upper house of Parliament), and Election Commission websites. The idea, he says, is “to bridge a gap between the common people and the functioning of government.”

Growing up in the free and open source software movement, Surve says he just feels that data should be publicly available. Open data for him is not necessarily political or economic in nature, but rather is about a choice—an option to look at the workings of government if and in whatever respect one wants. He believes that most of his peers share his predilection for openness in all forms.

While trying to create a mashup bringing together profiles of MPs with records of their participation in Parliament (for GovCheck), Surve was frustrated by the lack of vision in the Election Commission’s data sharing practices, which he felt were done just as an exercise with little forethought. “The commissioners do not understand how people might use the data even though they have accepted the idea they should make it available,” Surve says. They don’t get the “remixable form” because they are thinking about data only from their perspective as users.

After spending about a week parsing government data online in formats, Surve quickly tired of it, as even the websites are difficult to navigate. Moreover, many government websites resolved only to IP addresses (rather than a domain name), so that he was never entirely sure whether he was looking at official government data or not. In the course of familiarizing himself with the ways that India was providing data, Surve began to consider open data schemes elsewhere, especially in the U.S., and he decided to work on easing access to certain government data for other programmers by developing an API. Particularly, he decided to programmatically expose “all data around civic participation”, including state assembly elections in Maharashtra and profiles of MPs. As a result, he and a partner spent about two weeks manually scraping information from government websites, which he has compiled into a database with which the OpenCivic API interacts.

Although he has managed to scrape government websites for their data, Surve believes that the government should institute an open data policy under which departments would be required to provide whatever raw data they possess and can readily give from a technology standpoint. “We have the capacity. It’s not difficult and it doesn’t have to be that comprehensible. We should start off with something that we can readily give and keep adding to it,” he says. If data were easier to access, he suggests, more people would be developing projects like OpenCivic.

In fact, Surve suggests that open data in India might at first revolve around digitizing RTI requests, so that any time a person receives an answer to a request, either the government or the individual who made the request uploads the answer to a centralized public-access server.⁷⁸ But the true difficulty lies not so much in making data available, Surve suggests—although that is hard as well—as in making it viable. “Without end user applications, this data is almost useless,” he says, especially to the majority of Indians who are not very technologically savvy. In order to foster an ecosystem where government data is shared from its source and in which developers do interesting things with that data, Surve promotes the idea of competitions on top of data, whereby the government or other institutions might offer money, prizes or other incentives for the development of useful applications using government data. Most important, though, in light of all the barriers to open government data, is that the ecosystem is designed and developed as a whole and with a purpose. “Unless there’s a vision, I don’t think it is going to happen,” he says.

Forms of Access to Data

These different organizations have very different forms of access to governmental data. Some of these have had to collect data themselves, but with help from the government (Akshara Foun-

⁷⁵OpenCivic, <http://www.opencivic.in>

⁷⁶AskNeta, <http://www.askneta.com>

⁷⁷GovCheck, <http://www.govcheck.net>

⁷⁸This is happening to a small extent with some Information Commissioners putting up appeals and replies online.

dation), some have access to the data as they are helping the government (Mapunity), some others have taken data that the government has published online and has been scraped via a script (judgments in the case of IndianKanoon, and candidate data in case of the work done by Association for Democratic Reforms), some used connections within government as one of the routes of getting access to information (Empowering India), and yet others have turned to crowd-sourcing to collect the data (busroutes.in). In this regard, it is useful to keep in mind the redesign of information architecture that has been proposed by the UK Power of Information Taskforce.⁷⁹ Instead of information being separated into three layers of raw data + analysis + presentation, instead it makes sense to have more layers: raw data + open API/downloads + analysis + open API/downloads + presentation + interaction.

Challenges

There are many challenges that face any call for open government data in India. Even if tomorrow a policy dictated that all identified government data be put up online in a format such that it be machine-readable, reusable, and easily interpretable, such a policy would not be effective. Simply put, the infrastructure to deliver on such a policy does not currently exist in India. These are problems that affect information infrastructure and effective e-governance, not just 'open government data'. Beyond infrastructure and data collection issues there are other issues raised by the issue of openness and also that are particular for the Indian environment.

There are issues around cultural shifts necessary for open data to have a net positive effect that policy has to deal with. Issues of uneven power relations have to be taken into account to ensure that the elite don't take undue advantage of the marginalised by exploiting open data. Issues of privacy have to be studied carefully to ensure a balance between the benefits from opening up data and the risks of sharing personal or private data. Also, sufficient momentum needs to be developed within the government to push it towards open data.

First, there is a concern that, because certain information endows its guardians with power over those who lack access to it, many bureaucrats will feel threatened by open data; second, that some bureaucrats will resist open data due to a belief that people are either not responsible enough to acquire certain information or would seek to use it irresponsibly (as the media often do); and third, that bureaucrats are averse to any extra effort as would be involved in opening their data, and thus will practically neutralize any open data policy by finding ways to avoid such work even if given a mandate. There are some people, both within the government and in civil society, such as Chakshu Roy of Parliamentary Research Service, who don't believe that such fears are warranted—and that the more senior bureaucracy, at the very least, understands the importance and the need to move towards greater transparency and is slowly doing so. In general, though, on the specific question of having an open data policy mandate, most interviewed feel that a mandate would be useful, but would ultimately be insufficient to bring about transparency and accountability.

Currently, all government departments have websites where they disseminate information about their functions, contact details for officers, projects-specific information, etc., including annual reports providing information on activities and finances. Many of these reports, however, are image PDFs implicating a host of accessibility issues. Much of the information found on the web-

⁷⁹<http://powerofinformation.wordpress.com/2008/06/19/more-architecture/>

sites don't have any timestamps and are sometimes clearly dated. Moreover, even if updated annual reports are to be found, they usually contain no raw data, only processed information. For example, patent statistics are available from the Patent Office only in aggregate form in its annual reports⁸⁰; the raw data from which the statistics are compiled remains inaccessible. Dr. Govind, head of the E-Infrastructure Division in the Department of Information Technology, notes that "it may not be possible to put ... very detailed project proposals, outcomes and all" in departments' annual reports, but suggests that in large part information on "the major outcomes, major milestones, what is the direction and what is the vision of the department" are sufficient, as that is what people are looking for in an annual report. He also noted other problems with putting up data online: it is usually done only in English, thus even literate and Internet-connected persons may not have access to it unless they can read English.

Dr. Govind contends that most information that people want is already provided on the department websites, while admitting that most government data available online is not provided in formats conducive to analysis, but implies that there is not sufficient demand for the kind of sophisticated analyses encouraged by initiatives like the US and UK open data schemes. "Most of the websites are made for the common man [meaning] that the layman should be able to go and see [information] in a readable format," he says. Furthermore, a lot of data processing still occurs on paper and is only input into computers at a level beyond raw data, so even where there is demand for raw data current practices do not enable its electronic dissemination. There is also the problem, that "[Indian] society is basically an oral society [and is] not documentation driven." Slowly, however, he thinks that the government is moving towards computerized data processing at all levels of collection and analysis; the problem not yet properly addressed is that of universal standardization.

Beyond standardization, though, in many government offices—particularly at lower levels of government—computers are not much more than advanced photocopiers which also provide digital storage space for information primarily dealt with on paper.

Shailesh Gandhi has worked with the NIC to make his office paperless, but the custom-designed system extends only within his personal office and not to the CIC at large. He has, however, urged others to follow suit. The question, perhaps, is what policies—if not a mandate—might incentivize government offices to become truly paperless.

However, Dr. Govind believes that thinking of moving the government in particular directions in terms of 'incentives' is misguided. He points out that governments exist for the welfare of their citizens, and getting the government to move in any direction must be shown to be good for the citizens, and not in terms of 'incentives'. In other words, the issue is not forcing disclosure, which is already happening, but tailoring data collection and dissemination to the needs of citizens throughout society. In order to incentivize people to use data, he says, government must take a multifaceted approach focusing on creating awareness, propagating open standards, ensuring accessibility, and taking into account language barriers and varying levels of literacy. Mr. Aggarwal notes that the state government of Maharashtra, for instance, provided cash incentives to people who would undergo computer training. But the people merely took the cash without making an

⁸⁰<http://www.patentoffice.nic.in/>

effort to learn how to use computers, since there was no follow-up incentive (material or otherwise) for actually using computers at work. The lesson seems to be that in order to promote robust e-governance practices, both the government and citizens must take strides to learn new ways of interacting with their environments.

Beyond compliance with Section 4 and other issues related to automation of government processes and the resistance to such, there is also the issue of language barriers; websites which are not in the local language are relatively useless at lower levels of community. And even where information is available in the local language, most people—at least in villages—lack the awareness necessary to use and react to the information.

Data Unreliability

And when people do use the RTI, they are not guaranteed of the veracity of the information with which they are provided. For example, official statistics for murder rates in Mumbai from 1983 to 2007 collected by Shailesh Gandhi through an RTI request show that murders increased by 5000 per year from about 35,000 to about 40,000, while the population of Mumbai itself during the same increased by 30 per cent to 40 per cent. The discrepancy between the respective increases in murders and population suggests that one or both of the figures are unreliable, but there was no indication that the data is unreliable outside of the results of the RTI requests.

Murali Mohan, director of Sadhana—an NGO exploring alternative approaches to education in Andhra Pradesh—says that once his organization obtains data from the government, which can be difficult in itself, it often finds that the data provided on education is either false or misleading. He attributes this to a lack of incentives for government agents to provide quality information as well as a lack of civil society accountability programmes checking government data for veracity. Additionally, he says, opaque methodologies for data collection make it difficult to assess the quality of the education data that is available.

These reliability issues have been faced by most people interviewed from civil society. Thus, people must use the RTI not only to access information but also—to the extent that comparisons across datasets can be made—also to determine its reliability. And even where data is reliable, issues of disparate terminologies and methodologies across government and civil society call into question the actual value of public data.

This is an area where OGD can help out greatly. Hitherto finding out whether data is incorrect or not has been a difficult process involving requesting all the different data sets that one wishes to compare. However, if OGD becomes a reality, comparing datasets becomes a much easier task just by virtue of the fact that the first hurdle of getting data out is already crossed.

Semantic and System Interoperability

Semantic Interoperability

Many of the difficulties that are faced by groups like IndiaGoverns and Akshara while working with government data are because much of the data that they get from the government are not semantically interoperable. Different governance units might have been used in measuring related data, for instance. Or the same term might be used with different meanings in different departments' reports (or even the same department's in different points of time). Thus making sense of

the data becomes difficult if not impossible. According to the EU's Interoperable Delivery of European e-Government Services to Public Administrations, Business and Citizens (IDABC), semantic interoperability is concerned with ensuring that the precise meaning of exchanged information is understandable by any other application that was not initially developed for this purpose. One solution for semantic interoperability is setting and employing data and metadata standards. The Indian government is currently doing so for many categories through the e-Gov Standards Apex Body.

System Interoperability

An important question in considering data practices is whether they are vendor-driven, in which case there may be little or no incentive to formulate open standards governing the accessibility of data. While Mr. Aggarwal does not believe that open data in India is a vendor-driven process, he does assert that because the government currently has no open standards policy, many departments have engaged in standalone projects. "When you go in for stand alone projects there is always the danger of creating silos," he says. "Because if you are creating an application on your own without taking into consideration the issue of standards, then there may be an issue of interoperability. So the info available to you may not be then shareable with me because there [are] no mechanics [for sharing]." But as certain standards evolve and come into practice (not necessarily by the imposition of a central mandate) and departments begin to adhere to those standards, interoperability becomes less of an issue. Such standards are already developing—although current practices indicate that they may not represent what open standards do in the US and UK, for example—and most states are adhering to them.

Regardless of whether data practices are vendor-driven, there are issues not just of formulating open standards but of working around proprietary standards in use by certain government offices. The Income Tax Department, for example, allows people to file online for their tax returns, but only via a Microsoft Excel macro. In a country where many people cannot afford or do not have access to Microsoft Office products, the lack of ability to file tax returns using free and open source software presents a barrier—surmountable by filing at a common service centre, but there nevertheless—even to that portion of the population which does have Internet access.

There are thus many components to system non-interoperability:

- Lack of transparency and inter-departmental coordination in data collection
- Lack of good internal record-keeping practices
- Lack of interconnections between data sets by different departments, and cross-verification
- Lack of interoperability between the different formats in which data is published
- Bottlenecks in web publishing, especially due to not using content management systems and centralizing web publishing authority within a department. This results in delays and often in non-publication of information. Thus, one department will often not get to find out what other departments are doing, leading to different departments working in silos.

Some of these concerns are examined at greater length in a paper titled 'Selected Aspects of

Interoperability in One-stop Government Portal of India' by Rakhi Tripathi, et al. (2007),⁸¹ looking in particular the challenges in developing a one-stop portal for public-facing e-governance.

Cost of Data

It is often pointed out that most of the information that people seek access to via RTI, or data that people talk about in terms of OGD, are information and data that the government collects for its own internal purposes. In other words, whether or not people sought access to it, the information would have to be gathered, and thus no additional costs are incurred in the initial data collection itself (while some, though much lower, costs may arise in information collation and information disbursement).

In practice, in some cases, government information is available digitally but not free of charge. This is the case with information ranging from company records maintained by the Ministry of Corporate Affairs (the MCA21 MMP) to the eGazette of India, the electronic version of the official repository of all Central Government notifications, and the official source of all national laws.⁸² While some basic data from the National Sample Survey is available for free, any serious research requires people to pay.

Mr. Shankar Aggarwal says that the explanation in the latter case, and in similar cases, is that "At some point of time, because everything had to be put on a piece of paper," the government had to cover its costs. While costs associated with publishing information have decreased as the government has moved to automated processes, the mindset whereby people were charged for that information has not yet shifted in kind. "The government is by nature a little status quoist. It takes time to adopt new system, to adopt new ways and means of doing business. But I am certain that today we have come to a stage, which is basically a take off stage, where most of the services can be delivered electronically, and in another 2-3 years' time, you will see that majority of the services will be delivered only electronically," Mr. Aggarwal assuredly promises, pinning his optimism on the success of the Mission Mode Projects under the NeGP.

However, the National Knowledge Commission itself seems to have taken the view that e-governance cannot happen sustainably with free access, and envisions a system based on user fees. Recommendation 6 ("web services") of the NKC e-Governance Recommendations document reads: "To enforce standards and to keep the governance uniformly responsive and transparent, it is recommended that state governments use templates created by the Central Government to offer localized data and services in Indian languages. In this model, the private sector can invest in creation of access-infrastructure and building relevant business models for user-fee collection and its sharing across all stakeholders, to ensure sustainability and adaptation for future needs. This also implies that all public institutions will make sure that all public data is available on the web."

Mandakini Devasher of the Centre for Policy Research notes that the public-private partnership model might harm access, as seen in the case of MCA21, which was developed by a private software company. The e-Gazette portal was developed by NIC, and while it is currently not accessible without a fee, the government recently announced that this is soon to change. Additionally, other e-Gazette portals which NIC developed or helped develop (such as that of Bihar, and of Hi-

⁸¹http://www.iceg.net/2007/books/2/1_414_2.pdf

⁸²<http://egazette.nic.in/>

machal—a state which publishes an online-only Gazette) are accessible for free.

While providing web access might not incur too many additional costs on the government, running projects like Jaankari—a Bihar-government run project for providing RTI facilities through call centres—cost much more money, while reaching out to disadvantaged and digitally unconnected audiences. Thus sustainability is a larger factor there.

In this regard, Shailesh Gandhi pointed out that government e-governance budgets are extremely large and large costs such as that of data digitization, free access over the web, etc., are not really issues if the political will is there.

Scalability

There are very few large, sustainable groups working on issues of governance and technology. Even the governance-related NGOs that are venturing into technology don't usually envision it as a sustained effort, and limit themselves generally to providing analysis of governmental data on their website as an additional mode of information dissemination. Well-organized civil society organizations and civic hackers often don't end up working with one another. This can be rectified, though, as pointed out in the Recommendations section of the paper.

Indian Names

Dealing with the names of 1.2 billion people divided by language, religion, caste, ethnicity, etc. can raise some challenges that are peculiar to India. While some of these challenges can be at least partially overcome through the use of advanced technologies, others require substantial re-negotiation of social practices and power-relationships.

First, Indian names pose particular privacy problems, as it is often possible to accurately determine from a person's name her religion, place of origin, and caste, the last of which has in contemporary India infiltrated Christianity and, to a lesser extent, Islam. For example, Ambrose Pinto is most likely the name of a male, upper-caste Christian man from Goa. Since tax records and voters databases have been alleged to have been used in the past during incidents of communal violence,⁸³ open data projects seeking to uphold the privacy of ordinary citizens must by design be sensitive to name-related issues.

Second, name changes are very common and often unrecorded. For many, for instance dalits and Sikhs, getting rid of a caste-name is form of emancipation with political or religious overtones. Others, on advice of astrologers and numerologists, change the spelling of their names, modify existing names, etc.

Third, Indic scripts are generally phonetic, so it is usually easy to determine the pronunciation of names, with occasional idiosyncratic variations. The Roman script pronunciation, however, is not phonetic, so names are usually rendered in several variations. In addition, different generations of diaspora populations have also standardized different variations of renditions in Roman script—for example, the same name may be rendered as "Dixit" and "Dikshit." This makes it particularly difficult to do automated or manual quality control and maintenance on open data sets involving names. It also makes it more complex to combine different data sets which may each be based on

⁸³<http://www.hinduonnet.com/fline/fl12624/stories/20091204262410000.htm>

a different Indic script across multiple national, regional and local databases, and interferes with user expectations of search, sort and index functionality. The problem is compounded further as Indians become increasingly mobile. Transliteration technology is evolving, however, and perhaps this challenge maybe more adequately addressed as datasets grow in size.

Finally, the very concept of surnames/family names don't exist in many Indian cultures. Malayali Christians in the southern state of Kerala, for instance, have a tendency to mix the components of parent names to invent totally new names; adopt names of international public figures, professions or brands; adopt words of Indic or foreign languages either in original or modified form; etc. Some Sikhs in New Delhi have adopted their neighbourhood's name as part of their name.⁸⁴ It is also not always possible to tell the relationship between parent and child or siblings through the names of persons from certain Indian communities, because there is no standard modality for prefixing or suffixing the complete or abbreviated name of the parent.

While the UID system now calls for a given name and a surname (with a middle name being optional), many Indians, especially South Indians, don't follow a system of surnames (e.g., Parliamentarian E.M. Sudarsana Natchiappan, whose father's name was N.V. Madhavan and mother's name was Thangam). And as that example shows, many have more than three names. Many women either do without a surname, or adopt their husband's given name as their second name.

There is now a metadata and data standard for person identification that has been approved by the e-governance standards apex body.⁸⁵ How successful this is in solving the myriad problems associated with Indian names remains to be seen.

Indic Languages and Technology

India has Hindi and English as its two official languages. Apart from these, the Constitution recognizes 21 languages other than Hindi in the Eight Schedule. Additionally, there are hundreds of smaller languages and dialects, and as a consequence the business of state, district, block and village level administration is conducted in a wide range of languages across the country. The fractured nature of the Indic-computing market means that even large commercial players like Microsoft only provide support for 12 languages. Free software distributions like Ubuntu may provide additional support, but serious problems remain. In particular, problems centre on:

- Character encoding—Unicode adoption is only gradually happening. Most data in government offices with older hardware and operating systems is available only in font encoding. Font encoding does not guarantee search, sort, index or scrutability by search engines.
- Non-unicode standards—The Indian government has also developed standards such as ISCII and TSCII. Reasons for this include:
 - non-participation in Unicode;
 - unwillingness of Unicode to meeting demands of the language community;
 - lack of computing resources: for example, 8-bit TSCII encoding would be faster than 16-bit Unicode encoding;
 - legacy hardware and software; and

⁸⁴<http://www.hindu.com/mp/2003/10/30/stories/2003103000520300.htm>

⁸⁵http://egovstandards.gov.in/egscontent.2009-12-29.9641944881/at_download/file

- legacy data and technology, which make shifting to Unicode an expensive and disruptive proposition without any obvious internal benefits.
- Machine Translation—Even though translation engines have been developed by private corporations and government funded labs like CDAC for languages such as Bengali, Gujarati, Hindi, Malayalam, Marathi, Oriya and Urdu⁸⁶. However, the lack of dictionaries and corpus means that accuracy remains at unacceptable levels.
- Voice Integration—Voice recognition systems are increasing becoming available for limited vocabulary thanks to mobile value-added services leaders such as OnMobile. However, these are still unaffordable to most NGOs and government users. The situation with text-to-voice engines is slightly better given free software projects such as Gnome-speech. In India, voice integration is particularly important given the high penetration of telecommunications and that one in every two Indians is illiterate.
- Optical Character Recognition—Automated conversion of legacy data into electronic formats is easier in languages based on the Roman and other commonplace scripts because of the availability of robust OCR technologies. For Indic languages, OCR technologies are at a very rudimentary stage of development, creating a host of problems when it comes to digitizing government processes. The RTI, for example, has been successful in part because information officers must accept handwritten requests, which enable people who lack access to computers and/or the Internet to petition the government for information. However, neither the CIC nor other information offices can digitize many such petitions in machine-readable formats, because OCR does not support all of the various scripts in which requests may be submitted. Thus there is a disjunct between the mechanisms enabling access to knowledge under the RTI and the kinds of processes inherent in robust e-governance practices.

To begin dealing with problems related to translation and access to information, the NKC has recommended that India take steps to expand its translation industry, which the commission believes may potentially employ as many as half a million people. It has also recommended that the government create, maintain, and constantly update “a store-house of information on all aspects of translation involving Indian languages.” But the core problems of digitization and providing machine-readable data in local scripts and languages remain unsolved.

Technological Efficiency versus Access

One of the strongest arguments in favour of the RTI Act in India is that it is very simple to use. All one has to do is to pay ten rupees (generally) per application, and ask the questions, addressed to the correct PIO (Public Information Officer), providing contact details. In this, there is an interesting parallel with the way the Supreme Court of India's embrace of its "epistolary" jurisdiction, a trend championed especially by Justices Krishna Iyer and P.N. Bhagwati in the nineteen eighties, opened up the gates of justice to tens of thousands. By lowering the requirements of *locus standi* and allowing mere letters to be treated as writ petitions, the Supreme Court allowed vulnerable sections of society access to the highest court without overdue procedural complexity, and it also allowed for those who genuinely represent larger social interests (such as those pointing out violations of fundamental rights, and environmental degradation) to present a case before

⁸⁶http://pune.cdac.in/html/gist/research-areas/nlp_mt.aspx

the court.

Similarly, the RTI Act, by not prescribing any particular format for the RTI request, makes it as easy as sending a letter.⁸⁷ However, by striving for such simplicity, it runs counter to the direction in which advocates for open data would wish. Given that anyone can write an RTI request by hand, in any language, it becomes an impossible task to have machine-readable digital copies of these requests. This is borne out well by Mr. Shailesh Gandhi's efforts towards a paperless office show: the twin aims of digitization and allowing for access to information via a simple letter in any language are in conflict with each other. As Mr. Gandhi pointed out that no optical character recognition software available today could possibly render all the different letters he receives into a text-searchable database.

This problem, of course, isn't one of RTI alone. It is a more general problem. As many of the interviewed RTI activists asserted, delivery of information through the World Wide Web does not necessarily exclude the poor from accessing it, especially in light of intermediaries such as grass-roots NGOs.⁸⁸ However, delivery of information through other channels might in many cases increase visibility (as with blackboards outside Panchayat offices) as that information would retain its contextual relevance. And such contextually relevant delivery mechanisms would exclude the Internet without additional effort (and some would say additional redundancy). Thus, whether some digitization is merely redundancy or whether it actually adds value by creating efficiencies of search, retrieval, etc., must be judged. It must be noted here that not all redundancy is bad, as digital backups show—redundancy is often a desirable criterion in an information storage system. However, the costs of digitization must be commensurate to the benefits derived from it, for otherwise data backup is an endless process (should one keep a backup of the backup?).

Privacy

The Indian government, cognizant of the fact that securing growth requires heavy investment in electronic security and privacy frameworks, is currently considering new regulations for data protection. The Data Security Council of India (DSCI), set up by NASSCOM, a software industry body, to promote and help ensure the "trustworthiness of Indian companies as global sourcing service providers," is currently consulting with the government on these regulations. In addition to ensuring the economic value of data flows, which decreases as access to them is compromised, the DSCI is also concerned with protecting the individual's right to her personal information.

Vinayak Godse, director of data protection at the DSCI, notes that in India, people are much freer with their personal information than their counterparts in Western countries, although he believes this is gradually changing due to people gradually leaving behind living with large, extended families in favour of nuclear families where they have more privacy. He believes people in urban areas especially are becoming more concerned with privacy and the security of their personal data as consumer and other transactions increasingly take place online with the propagation of e-governance and banking applications. A problem, though, is that as e-government applications in

⁸⁷ It must be noted that some RTI activists argue that a standard form would help, because then the possibility of incomplete applications would decrease. This is also noted in the PWC report prepared for the Department of Personnel & Training. However, even if a standard form is prescribed by any authority, an application that does not conform to that standard form cannot be rejected due to the Act itself not prescribing any such form.

⁸⁸ Interviews with Mandakini Devasher Surie of Accountability Initiative and Venkatesh Nayak of Commonwealth Human Rights Initiative.

particular seek to integrate people at all levels of society without respect to whether they are computer literate or otherwise technology savvy, many users of such applications do not understand the implications for their privacy of certain actions they may take online. Thus as Internet use expands across India and deepens into society, individuals may be risking their personal privacy or foregoing it altogether without ever considering what they are doing—especially considering that many e-governance initiatives were not developed with privacy in mind.

The solution to this problem, says Godse, is that any comprehensive data system must incorporate privacy into its design, so that it will be there whether or not people care about their rights or even realize that they might want to care; also existing e-governance initiatives must be revisited from the perspective of protecting personal information. Certain privacy principles should simply be adhered to from the outset, and where systems already exist, they should be updated to adhere to these principles. This, of course, calls into question what these principles should be, how much weight they should carry, and whether they should carry that same weight in all places and under all circumstances. With regard to these latter aspects, Godse thinks that in India, some privacy principles will vary from state to state, reflecting the cultural variances in a society as large and diverse as India—so developed states like Maharashtra, for example, will have a more robust privacy regime than less developed states like Bihar or Orissa. But in any case there must be a base set of principles adhered to throughout the country and at all levels of governance. Probably the biggest challenge in ensuring privacy and data security lies in creating awareness of such issues among end users, which Godse suggests might be accomplished through mass media campaigns.

While awareness-building might work as a solution for clear-cut cases where privacy needs to be safeguarded, the truth of the matter is that privacy is often not a clear-cut issue at all. For instance, during the last general elections, electoral rolls were made available for online searches by many Election Commission websites, as did Google. Given that many people would have to depend on others to check their registration online, it was probably a good idea to anyone to search for any name online. However, the details provided included name, age, and address. Thus, one could potentially find out, from the comfort of one's chair, how many people above the age of 18 resided in any house, how many of them were elderly people, as well as be able to make reasonable guesses as to their religious and caste identities. Should this convenience provided in a country where assistance with Internet usage is a reality be considered an invasion of privacy? Anonymization of this data set would make it useless, because to verify the correctness of the data (and to find out information like what one's polling booth is), one needs to be able to correlate the name with the age and the address.

Even if anonymization happens, there are many situations where privacy concerns would still remain. R. Siva Kumar of NSDI cited an example of the Health Ministry being a bit reluctant to make openly available important information on disease-tracking that would be of great use to researcher, health NGOs, etc. Their reason was grounded in a concern for privacy. Villages in which there are greater concentrations of HIV-positive people will be discriminated against even if specific individuals are not identified. While individual-level concerns may be taken care of by anonymization, community-level concerns are much more difficult to address.

In India privacy may also be seen (though not legally) as something that a community considers important, whereas in the West it is almost strictly interpreted as a right vested in an individual *against* community and the state.⁸⁹ The community at large takes questions of “shame” very seriously (leading even to murders in the name of community and family honour), and the publicness of an affair is a problem rather than the affair itself. And at any rate, there is a trade-off between the usefulness of information and privacy. Anonymized location data of HIV-positive patients would help in spotting patterns—like truck routes being particularly vulnerable—however those might also raise privacy concerns.

It is a given that privacy as a concept is closely related to data protection. There does exist data protection legislation in India; the Information Technology Act of 2000 addressed the need for some legal protections but not in an exclusive manner.⁹⁰ The Act provides for civil liability in if the privacy of data is ever violated through theft, it also protects databases and protects the data itself from being altered without permission. *“The Act is not limited to protecting computer source code only, but it also safeguards data and computer databases; and similarly section 66 [Hacking with Computer System] covers cyber offences related to (a) Illegal access, (b) Illegal interception, (c) Data interference, (d) System interference, (e) Misuse of devices, etc. Page 10 of 11”*

However, this isn’t a comprehensive policy regarding privacy. *“There is no an actual legal framework in the form of Data Protection Authority, data quality and proportionality, data transparency etc. which properly addresses and covers data protection issues in accordance with the principles of the EU Directive, OECD Guidelines or Safe Harbor Principles. Accordingly, even if the new proposed amendments to the Information Technology Act, 2000 were adopted, India would still lack a real legal framework for data protection and privacy.”⁹¹*

In order for e-governance and open data to be successful people need to feel secure that their information is safe when they participate. And there have been statutory enactments in the west that have given broader protection than the IT Act currently gives. Granted the IT Act is being updated and it will be interested to follow what changes are being added for privacy. For example the UK’s Data Protection Act, 1998, has incorporated eight data protection principles that the Data Controllers apply to personal data when it is being processed.⁹²

The following are the eight principles of personal data according to the UK law:

Personal Data : Eight Principles⁹³

Personal data shall :

- 1. be obtained & processed fairly and lawfully*
- 2. be held only for lawful purposes, which are described in the register entry*

⁸⁹In *Mr. X v. Hospital Z*, a 1998 Supreme Court case on whether the hospital was wrong in telling Mr. X’s fiancée about his being HIV-positive, the court held: “Having regard to the fact that the appellant was found to be HIV(+), its disclosure would not be violate of either the rule of confidentiality of the appellant’s Right of Privacy as Ms. ‘W’ with whom the appellant was likely to be married was saved in time by such disclosure, or else, she took would have been infected with the dreadful disease if marriage had taken place and consummated.”

⁹⁰ ‘White Paper on Privacy Protection in India’ by Mr. Vakul Sharma (<http://www.iamai.in/Upload/Istandard/White%20Paper%20on%20Privacy.%202007.pdf>)

⁹¹ Ibid

⁹² ‘Law and E-Governance’ by Vakul Sharma (http://www.mgsipap.org/egov-conference/Post_Conference/conf_Papers/Papers/Vakul_Sharma.pdf/)

⁹³ <http://www.legislation.gov.uk/ukpga/1998/29/contents#pt6-pb2-11g55>

3. *be used or disclosed only for lawful or compatible purposes*
4. *be adequate, relevant and not excessive in relation to the purpose for which they are held*
5. *be accurate and, where necessary, kept up to date*
6. *be held no longer than is necessary for the purpose for which they are held*
7. *be accessible to individuals it concerns, who may, where appropriate, correct or erase it*
8. *be surrounded by proper security*

It is clear that not only technological solutions are needed but also a real legal base of protection for people's data. This would make India's e-governance and open data schemes more secure and give people more confidence to use them.

Elite Capture of Transparency

One of the first research projects on locating the politics of open data governance structures outside the narrow administrative narratives was conducted by Dr. Solomon Benjamin, R. Bhuvaneshwari, P. Rajan and Manjunath. In an attempt to explore the ways in which existing accounts of e-government and digitization of data gloss over the role that different stakeholders play in implementing an efficient system of open government data management, the research concentrated on the first (and professedly successful) project called Bhoomi.

Bhoomi is a project that was launched by the Government of Karnataka to digitize land records and titles which enable the farmers and land-owners in rural districts of Karnataka, to prove their ownership over their land—a process that enabled them for several government schemes and subsidies and also other public services. The project was a pilot that has won much acclaim from around the world for the way in which public data which was hitherto unavailable except through bureaucratic processes and bribes, was made available to the public. The Bhoomi project sought to empower farmers by offering them their land record data through a massive infrastructure of Citizen Service Centres which were established at the village level through various public-private partnerships. The project has also been heralded as the model project to be established across the country.

However, the ethnographic research conducted by Dr. Benjamin and his team shows that the digitization of land records led to increased corruption, substantially increased bribes and time taken for land transactions. In their report they say, "Before Bhoomi, corruption did exist but was very less. Now, with the Bhoomi program centralizing land management and providing open access to land records, corruption is cumulative at various levels, and of a much higher amount." They trace these problems to a flawed blueprint that did not take into account:

1. Technology literacy—Many farmers were print literate and hence able to negotiate their way through existing processes, became unconditionally dependent on the CSC operator in the villages.
2. Economy Ecology — Despite the preventive technical measures such as the FIFO (First In, First Out) transaction model, bribery remained prevalent across all administrative and technical

structures due to lack of orientation for the people actually working on administering and managing the project, who only saw this as a way of collecting larger bribes. The beneficiaries themselves had no control or ownership over the technologies which were designed to give them access to their data.

3. The nature of Public Private Partnerships — A lack of transparency about the nature of the PPP itself led to severe exploitation where the big businesses of real estate development and IT, established monopolies over public land acquisition through their lobbying powers.

4. Lack of surrounding infrastructure — The big businesses and mega planning agencies were able to use the centralized land records for their benefits because the infrastructure for protecting the data was not established. The availability of the open citizens' data led to a compromise of their rights, far countering the benefits offered by the system.

The report tries to offer what it believes are concrete suggestions about what e-governance projects such as Bhoomi need to be sensitive about:

1. The projects need to be aware of the political economy of the fields they penetrate rather than getting lost in the techno-managerial features.

2. Effective processes within the field are societally evolved and reflect the consolidation of political and economic claims by a variety of groups including the poor. The project should ensure that these diverse claims and forms of entitlement are not erased in the new narrow frameworks which are shaped by available technologies.

3. The projects need to be aware of nuances of ownership and value. Just making available of the data to the citizens is not enough. Safeguards to protect the use of the data, training and orientation for all the staff involved in the everyday practice of the project are required. Strict guidelines and legislation needs to be in place to protect the data and the rights of the citizens.

4. Scalability needs to be replaced by community owned sustainability. Similarly, other corporate process based keywords like transparency, efficiency and best practices, need to be replaced by conceptually more rigorous terms that reflect the uneven terrain of power and control that the governance embodies.

While the analysis conducted by Dr. Benjamin and his team points to a real problem, the suggestions they offer are not necessarily practicable. It is important to realize that the problem is not one that can truly be solved within the framework of transparency and e-governance. Adding more transparency might not solve the problem, nor will reducing transparency. And it is the same with e-governance. Regardless of the type of project, important factors such as differential power relations need to be accounted for, and by doing so unintended consequences must be sought to be minimized. Even the Mahatma Gandhi National Rural Employment Guarantee Scheme has been criticised along similar lines, by pointing out that sometimes after the completion of a public work by the poor under the scheme, the wealthier sections of society appropriate it.⁹⁴

⁹⁴Disa Sjoblom and John Farrington, 'The Indian National Rural Employment Guarantee Act: Will it reduce poverty and boost the economy?', ODI Project Briefings 7 (January 2008), available at <http://www.odi.org.uk/resources/download/440.pdf>.

Problems such as the press misconstruing data (which was often cited as a problem, both by bureaucrats like Dr. Govind, as well as by persons from civil society organizations that aid access to data, like Chakshu Roy of the Parliamentary Research Service) can be solved within the frame of a transparency project by providing more detailed annotations, having a query helpline, etc., since they are problems relating to data interpretation. However, exploitation due to power imbalances will *necessarily* have to be remedied outside the frame of the project because the problem there is *use the data are put to*, which happens outside the frame of the project. While interpretational use of the data is something that can be addressed through transparency mechanisms, use of the data to do something else (in this case, buy land) is not. For that, the protection must be provided elsewhere (at the point of land purchase, for instance) and not at the level of the transparency project. After all, the point of Anatole France's famous quote about the majestic equality of the law⁹⁵ was not to do away with rule of the law, but to imbue it with a sense of justice and fairness.

Decentralization

Decentralization of power and fiscal responsibility could help empower local bodies to collect information because they will be responsible for services and in order to be accountable to citizens and track their own performance they will have to keep information more organized and complete than what is currently being done.

The 73rd and 74th amendments to the Indian Constitution set a starting point for decentralization to the panchayat and urban levels. However, critics say that the Amendments didn't go far enough and that there is still a great deal of power at the central and state levels. The main issue is that the State bodies are responsible for decentralizing and that unless the Chief Minister is particularly motivated there won't be a distribution of significant power. (http://www.indg.in/social-sector/nird/decentralisation_in_india.pdf)

Right now the Panchayats and urban areas of India are in an accountability haze. The functions that the local governing body has complete control over are limited and so is the budget that they have discretion over. Even things like the City Development Plan has to be approved by the State and Centre before it can be implemented by the city. A large part of the budget and programs are not created by the local government but by the Centre. This leaves the local government unable to control every aspect of their area. This means that even though a Panchayat council or mayor are elected they don't have much say in what happens in the area or control to change policy as they see fit. This has a direct relationship with how local areas collect information or how they do not collect information.

If a local government has no say in how a policy is formed or changed then there is no motivation to see if it is working or how it is being implemented. This causes a gap in information collection and how accurate the life of a Panchayat or city is being recorded. This tends to make information related to specific programs more robust (but barely) than data on everything that happens in a local region even if it not related to a national program.

Decentralizing power and financial responsibility would allow a closer relationship with citizens which could create motivation to collect data. "Decentralization doesn't necessarily improve the data management or collection skills of local governments in and of itself--effective data management and collection is more often seen as a key input in the decentralization program in order to achieve the objectives of decentralization i.e. more responsive, responsible and accountable government" explains Robert Boothe World Bank Public Sector Specialist. Decentralization creates a situation where people are closer to government and will require better information to create effective governance.

⁹⁵ "The law, in its majestic equality, forbids rich and poor alike to sleep under bridges, to beg in the streets, and to steal bread." Anatole France, *Le Lys Rouge* ch. 7 (1894).

Observations and Recommendations

General Observations

- Most people in India, even those who are active in issues of transparency, have never heard of “Open Data” or “Open Government Data”. Only a few civic hackers have. Despite this it might make more sense, for reasons given below, to build the movement for OGD out of the transparency movement (i.e., the RTI movement) rather than one focussed on innovations made possible by civic hacking. Most interviewees felt that an open government data mandate through policy would be useful, but insufficient without an effective system in place, for the promotion of transparency and accountability.

- Open Government Data in India should not be envisioned as it has been envisioned so far in other countries. The primary reason for this, as discussed below, is that Indian civil society is currently not configured around technology, and a developer-centric approach to OGD might not work as well, especially in terms of reaching out to the masses. Thus, civil society’s capacity to use ‘raw data’ must play necessarily a part in how the idea of OGD gets articulated. And just as promoting RTI awareness is mandated as part of the Act,⁹⁶ similarly, capacity to use OGD must be aided by the government.

- The existing provisions in the RTI Act and the National Knowledge Commission’s recommendations on e-Governance should form a basis for pushing for OGD in India.

- The National Knowledge Commission’s recommendations lay the basis for the informatics, organizational, software and network infrastructure that are required in order to provide open data. Very importantly, it sets out a roadmap for the reconfiguration of governance processes to account for ICT. This would be essential to ensure the sustainability in practice of any open data mandate. Thus, it is important to note that all the problems of e-governance in general are also problems for open government data.

- It also calls for online provision of data held by the government to the public as a means of ensuring democratic participation and governance. That document, while on e-governance, could easily serve as a blueprint for open government data as well, especially since it seeks to put in place infrastructure to make governmental data reach the rural poor as well.

- Section 4 of the RTI Act already provides for proactive disclosure. Any OGD-focussed policy should build on that. OGD might help bridge the gap between the individualised transparency that reactive RTI disclosure provides and the mass transparency that proactive disclosure could provide, and which is necessary to enable larger accountability.

- The benefits of OGD for governments themselves must be highlighted. For instance, inconsistencies in data will be made easier to spot if the information is made available freely, for many people to inspect. If this leads to more reliable data, that is to the benefit of the the government. Applications created by citizens may prove to be very useful for the government itself.

- While large-scale administrative reforms, especially in terms of improving the governmental

⁹⁶The Act requires appropriate governments to: “develop and organise educational programmes to advance the understanding of the public, in particular of disadvantaged communities as to how to exercise the rights contemplated under this Act” .

information processes, infrastructure, and interoperability, are needed if OGD is not to be as unsuccessful as current compliance with Section 4 requirements of the RTI Act. Without fixing the system, a sustained change cannot be brought about. Keeping OGD as a goal will help achieve these changes, and thus streamline governmental functioning. If data is easy for citizens to access, it will also become easy for disparate departments of government through its various tiers (Central, State, Panchayat) to access. Thus, the benefits of OGD are also benefits of e-governance.

- While set-ups like Common Service Centres with operators as intermediaries can be seen as reducing privacy and as disempowering when the operators are viewed as middlemen (and a potential point of corruption), more people view them as empowering for the masses who are illiterate or barely literate, than otherwise.⁹⁷

- Because of a relative lack of legacy hardware in India, it is in a position to leapfrog over ageing technologies. Relative technological backwardness thus provides a general advantage to many developing countries over developed countries.

Recommendations

Helping Governmental Policy Formulation and Execution

There is already some momentum on an open governmental data in India, which can be seen with the establishment of the plan for a national Public Information Infrastructure, and the efforts by the National Spatial Data Infrastructure Division of the Department of Science and Technology to study open data projects in other countries and move towards opening of government data while learning from the successes and failures of others. Thus, it would make sense to aid the government in coming up with best practices by looking at case studies elsewhere, and by providing research in India both on the data held by the government as well as the kinds of data that civil society organizations would be interested in working with. This would help the government set its policies and its implementation priorities.

Conceptualizing Open Data in India

The meaning of 'open government data' and the purposes they serve will have to be re-examined from an Indian perspective. The reasons that work well in the United States and the United Kingdom may not work well in India. We also have to be very careful about how we imagine the end users of open government data. Do we visualize open data as being for the benefit of individual middle class citizens by helping them consuming the (processed) data themselves (with bus routes, for instance), or do we visualize them as being for the benefit of the poor, and thus target target NGOs? Do we visualize them as being hackers or as laypersons?

All these questions of conceptualization are extremely important. While none of the above categories are either-or situations where one possibility will negate the other, still the choices make a big difference. Thus, while open government data could cater to both middle class concerns as well as those of the poor, which one a campaign focusses on will make a difference both

⁹⁷Apart from civil servants and government officials, many others such Venkatesh Nayak of the Commonwealth Human Rights Initiative, view them as positive. Venkatesh Nayak, for instance, cited the example of CSC deployment in Bangladesh as a very positive success story, worthy of emulation.

to get the idea of open data accepted amongst policymakers,⁹⁸ will affect the prioritization of datasets, the languages in which the data is made available, etc. It would make a difference to whether the government itself will have to invest in data visualization, sense-making of the data and data interaction, or whether they can choose to rely on citizen intermediaries to do so for the larger public. If the latter is the case, then the system will necessarily have to be built such that the government facilitates these third parties and aids the public in reaching these third-party applications.

Focus on Use and Benefit, Instead of Access & Potential for Reuse

OGD has generally been thought of in terms of access coupled with the legal and technical potential for reuse and sharing. While data mashing and private-sector information products are important goals whose viability must be ensured by any OGD policy that the government considers, the OGD policy must also, perhaps more importantly, ensure that it not only allows for data mashing, but facilitates and encourages it. In some cases, the government itself must create the applications that show potential uses. This would be in terms of a 'reference' for others (the way governments sometimes produce or commission reference implementations for standards), or as a means of ensuring relevance for those who fall on the wrong side of the digital divide (for instance training CSC operators to use these, or by running call centres the way the Bihar government did as part of the Jaankari project, etc.). The duty of the government cannot end merely at providing information, but must extend to making that available in such a form that facilitates analysis and enhances offline usability (for instance in a format that can easily be reproduced on a panchayat blackboard).

Further, governments must not focus only on the use and benefits that accrue to citizens, but also to itself. Many applications of the data are very relevant for the government itself. For instance, Arun Ganesh took the trouble to independently create (through crowd-sourcing) a map of Chennai bus routes because the concerned municipality department was not providing help. After he finished the task, he (who is also a student at the National Institute of Design in Bangalore) offered to provide the individually designed maps for different bus stops free of cost (where currently none exist). Instead of making the most of this opportunity, the municipal government let it slip after coming to a conclusion that putting up direction maps at bus stops blocks the flow of pedestrians.

Interestingly enough, the best online maps to be found of the Delhi metro are not on the Delhi metro website (where an ancient map is to be found), but on Wikipedia. The SVG image map on Wikipedia is openly licensed, and made by volunteers. The government could (and should) choose to make optimal use of the fruits of data availability.

In this regard governmental attitude must change, and citizen participation must be welcomed if true participatory democracy is to be realized.

Leveraging the RTI Movement

The RTI movement, which is very mature in India with many internationally recognized social

⁹⁸There are indications that focussing on 'the common man', and the disadvantaged and marginalised, would help make a stronger case than one for middle class considerations, and for social entrepreneurs and technologists and hobbyists.

activists often with good ties in government, would prove to be an ideal partner for any movement for open government data. While the priorities and demands of Western open government data activists and traditional transparency activists sometimes do not match (e.g., should the government be providing APIs to an XML datastore, or easier-to-use annotated spreadsheets?), that is not the case in India simply because the equivalent of the 'open government data activists' barely exist here. Thus the relationship between civic hackers and social activists is yet to be set, and leveraging the pull of one (the social activists) for the benefit of both would be the best way forward. Indeed, civic hackers are not in a position to influence policy change while RTI activists are.

Thus it would make sense for donors and policymakers looking to gain larger support for their promotion of open government data to contact groups like the National Campaign for People's Right to Information, and individuals like Aruna Roy, Nikhil Dey, Arvind Kejriwal, Shailesh Gandhi, Shekhar Singh, Venkatesh Nayak, Harsh Mander and other such stalwarts of the RTI movement.⁹⁹

Leveraging the FOSS and Open Content Communities

Civic hackers in the Western mould are the exception in India, rather than the rule. There are very few, and many of them proved to be the most knowledgeable about the open government data movement, and recent ideas such as "Government 2.0". Interestingly, almost without exception, the civic hackers we've identified in India are also advocates of free / open source software (FOSS). The individuals behind groups and projects like Mapunity, Mahiti, Busroutes.in, Indian Kanoon, iJanaagraha, OpenCivic, GovCheck, are all strong advocates of FOSS.¹⁰⁰ However, the idea of OGD has not really spread much in the various FOSS communities in India. These communities, which are formed on the basis of technology (languages like Python, platforms like Android, etc.), and geography (groups exist in a large number of cities), are highly technically competent and motivated (often with learning and fun as motivations). Hence, it might be possible to leverage these communities in India to contribute with code and help extend existing projects, and create new projects.

Many civic hackers are to be found in different 'open content' communities as well, specifically those around Open Street Maps, an openly-licensed volunteer-made map platform, and Wikipedia, an openly-licensed volunteer-made encyclopaedia. Both these projects have dedicated groups in India (groups like Geohackers and Mapunity for the former, and the Wikipedia India Chapter, and the newly-launched Wikimedia Foundation office in India to guide the latter), and are people who would both benefit from governmental data (reusable GIS information, scrapeable statistics datatypes) and would help take the raw data to other developers as well the general public.¹⁰¹

⁹⁹NCPRI and MKSS both signed on to the Civil Society Representation made by the Centre for Internet and Society to the Standing Committee of the Rajya Sabha looking at the Copyright Amendment Bill. Specifically, they both supported the recommendation made in the document to have a much wider exception for government copyright, since strong protection of government copyright harms the right to information.

¹⁰⁰As an interesting aside: the tech community has generally been very male-dominated and reflecting this the people who came for a civic hacking workshop organized by CIS with the UK government's Foreign Office / Cabinet Office Team for Digital Engagement and Google India, were also mostly male, as they represented the technologist side of their organizations. This is in contrast to the RTI groups where gender balance was maintained to a much greater degree.

¹⁰¹A good instance of this would be the work of Arun Ganesh. His site Busroutes.in uses mapping data from Open Street Maps. Additionally, he is a prolific contributor of India-related maps (including the most up-to-date route maps available of the Delhi metro) to Wikipedia.

Engagement with these communities and that of transparency organizations should happen both online through different mailing lists, as well as through face-to-face events like 'hackfests' in different cities that help members of a community meet interested peers as well as those from other communities, thus providing for cross-dialogue between those who can help with policy, those who consume data, and those who can play the role of technological intermediaries. It must also be noted that while we talk of different communities, there are no strict boundaries between them, and a good amount of overlap is not only possible but is also inevitable.

Annexure I: Survey of Organizations and Projects

Despite the lack of an open data policy in India, there are still some technologists working on open government data-based applications. However, the numbers are not even comparable to the number of organizations dedicated to RTI activism and governance issues. Technology is sometimes enlisted to aid transparency groups. The work of groups such as IndiaGoverns, Accountability Initiative, ADR, Janaagraha, SmartVote, etc., testify to this. It is less common to see technology-focussed organizations (of which there are a much lower number) be involved in transparency issues. Mahiti and Mapunity, which consist mostly of software engineers and coders, were both envisioned as social enterprises. Mahiti was spun-off from the technology wing of a larger NGO called Samvada. Thus, such groups straddle a middle ground as social-focussed technology groups, but having such as focus from their very inception. Examples of civic hacking such as IndianKanoon.org, OpenCivic.in and Busroutes.in are examples of the very limited third group: computer geeks getting involved in social projects involving government data. They are also usually individuals without larger organizational backing.

The first group—transparency groups using technology, often incidentally (e.g., Accountability Initiative at the Centre for Policy Research), though sometimes (e.g., Janaagraha and iJanaagraha, SmartVote, etc.) in a much more involved way—is the largest, while the second and the third groups are equally small. This might be one of the important differences between the Western OGD movement and the as-such-unlabelled OGD movement in India: the second and the third groups seem to be much smaller in India.

Most of the descriptions given below have been sourced from the websites of each of these organizations and projects.

Akshara Foundation is a Bangalore-based public charitable trust with the mission to ensure, that every child is in school and learning well. Established in the year 2000, Akshara Foundation has a range of programmes that provide, multiple solutions for universalizing elementary education. The Karnataka Learning Partnership was formed as a framework for non-profits, corporations, academic institutions, and citizens to get involved in improving government schools in Karnataka. It is a public space where all citizens can contribute to the cause of ensuring better schools and schooling for children.¹⁰²

Association for Democratic Reforms (ADR) is a non-partisan and a non-governmental organization. It was founded on August 1, 1999 by a group of professors from the Indian Institute of Management (IIM), Ahmedabad and National Institute of Design and some alumni of IIM to work towards strengthening democracy and governance in India by focusing on fair and transparent electoral processes.¹⁰³

Busroutes.in is a project founded by Arun Ganesh, which seeks to make publicly accessible maps of bus routes (currently in Chennai) available. Since trying to get the data required to plot the maps online using OpenStreetMaps was not forthcoming from the governmental department,

¹⁰²<http://klp.org.in/>

¹⁰³<http://www.adrindia.org/>

a crowd-sourcing approach was tried out. Within days, a workable map had been created, which could be queried using point of departure and destination.

The **Central Information Commission** is the final appellate body under the Right to Information Act. It is headed by a Chief Information Commissioner, and is the body responsible for overseeing compliance with the Act.¹⁰⁴

The **Centre for Budget and Governance Accountability** (CBGA) is an attempt to promote transparent, accountable and participatory governance, and a people-centred perspective in preparation and implementation of budgets. CBGA came into being at the initiative of a number of concerned individuals and organizations, many of which were already engaged with budget work: Development Initiative for Social and Human Action (DISHA) in Ahmedabad, Centre for Budget Studies (CBS) at Samarthan in Mumbai, Centre for Budget and Policy Studies (CBPS) in Bangalore, Social Watch Tamil Nadu in Chennai, and National Centre for Advocacy Studies (NCAS) in Pune. Many of them felt that what was conspicuous by its absence was an organization which would do focused work on the Central Government's budget and policies since in a federal structure, allocations made by the Central Government become very crucial as well. Created in 2002, as a programme of the National Centre for Advocacy Studies (NCAS), Pune, CBGA has now evolved as an independent organisation, analysing Union Budget and public policies from the perspective of the poor and the marginalised.¹⁰⁵

The **Centre for Policy Research** (CPR) is a non-profit, non-partisan and autonomous research institution and one of India's premier think tanks in public policy. Established in 1973 and located in New Delhi, it is one of 27 national social science research institutes recognized by the Indian Council of Social Science Research (ICSSR), the apex advisory body of the Government of India for the promotion of research in the social sciences. The objectives of CPR are to develop substantive policy options on matters relevant to the Indian polity, economy and society; to provide advisory services to governments, public bodies and other institutions; and to disseminate information on policy issues through various channels. The governing board of CPR consists of various public figures from Indian government, academia, and industry. Pratap Bhanu Mehta, a prominent political scientist and public commentator, is President and Chief Executive of the institution.¹⁰⁶

The **Commonwealth Human Rights Initiative** is an independent, non-partisan, international non-governmental organisation, mandated to ensure the practical realisation of human rights in the countries of the Commonwealth. They work extensively on the Right to Information Act as part of their Access to Information Program.¹⁰⁷

Empowering India is an election-related project run by the **Liberty Institute**, which is an independent think tank dedicated to empowering the people by harnessing the power of the market. It seeks to build understanding and appreciation of the four institutional pillars of a free society – Individual Rights, Rule of Law, Limited Government and Free Market. The Institute undertakes a number of activities, among them research and advocacy on public policy issues. It organ-

¹⁰⁴<http://www.cic.gov.in>

¹⁰⁵<http://www.cbgaindia.org>

¹⁰⁶<http://www.cprindia.org>

¹⁰⁷<http://www.humanrightsinitiative.org>

ises conferences and seminars, and has a publications programme. The Institute is working on a range of public policy issues, including economic development and trade policy, energy policy and environmental quality, education and health policies, democracy and governance reforms, intellectual property rights and innovation, among others.¹⁰⁸

Kiirti, in Sanskrit, means report or reputation. It is a platform to enable effective governance by promoting awareness and citizen engagement. It allows government, non-government and civic organizations to engage with citizens easily through phone, SMS, e-mail, and the web.¹⁰⁹

The **India Water Portal** (IWP), set up by Arghyam Trust, a public charitable foundation founded by Rohini Nilekani. Arghyam, a Sanskrit word meaning “offering,” has been active in the water and sanitation sectors in India since 2005. A detailed case study on the India Water Portal follows this summary of the NKC’s portals.¹¹⁰

The **India Energy Portal** (IEngP), run by The Energy and Resources Institute (TERI), an organization established in 1974 to find solutions to environmental problems caused by “the gradual depletion of the earth’s finite energy resources which are largely non-renewable and on account of the existing methods of their use which are polluting.” The IEngP features a knowledge bank providing statistics offering “insight into Indian as well as global energy scenarios in terms of resources, demand, supply, and installations; among others.” Currently, the portal hosts data like annual production of solar cells and modules; year-wise solar water-heating system installations; production of primary energy resources of conventional energy in India; comparison of intensity of India with mature and emerging economies; and estimated national energy demand, as well as detailed information in areas including waste, power, and energy maps. Unfortunately, information is provided mostly in HTML, and some links do not work.

The **India Environmental Portal**¹¹¹ (IEnvP) is run by the Centre for Science and Environment (CSE), an NGO founded in 1980 by environmental activist Anil Agarwal that “researches into, lobbies for, and communicates the urgency of development that is sustainable and equitable.”¹¹² The IEnvP, which runs on the open-source software Drupal, seeks “to put together a one-stop shop of all that you want to know about environment and development issues,” offering a centralized repository of environmental information from “research institutions, government bodies, NGOs, universities, the mass media, and experts, among others.”¹¹³ Although the portal contains a “Data and Statistics” section, most information is in the form of HTML reports and articles initially published in Down to Earth, a fortnightly publication of the CSE.¹¹⁴

The **India Biodiversity Portal** (IBP), set up by the Ashoka Trust for Research in Ecology and the Environment (ATREE), an NGO promoting “*socially just* environmental conservation and sustainable development by generating rigorous interdisciplinary knowledge that engages actively with academia, policy makers, practitioners, activists, students and wider public audiences.”¹¹⁵ The

¹⁰⁸<http://empoweringindia.org>

¹⁰⁹<http://www.kiirti.org>

¹¹⁰<http://www.indiawaterportal.org>

¹¹¹<http://www.indiaenvironmentportal.org.in/>

¹¹²<http://www.cseindia.org/node/214>

¹¹³<http://www.indiaenvironmentportal.org.in/content/about-us>

¹¹⁴<http://www.downtoearth.org.in/>

¹¹⁵<http://www.atree.org/purpose>

IBP is intended to “provide information on all aspects of biodiversity in India,”¹¹⁶ which it makes available via mapping technologies and geographic visualization software developed in-house that it deploys in conjunction with Drupal. The IBP’s generic “map-based web application platform” uses available Internet map servers like <http://maps.google.com> and <http://maps.yahoo.com> to render geographical visualizations and “provides the ability to overlay custom geographical content on the base maps.” It also allows users to upload “geo-referenced location-based information” to the platform, which then aggregates and serves that data to other interested users.¹¹⁷ The IBP’s platform has been developed as open source, and the India Water Portal is considering its use on their own website.

Janaagraha is a Bangalore-based civic and democracy NGO that works with citizens and government to improve the quality of life in India’s urban areas. The term “quality of life” has two separate aspects, distinct and inter-related: the first is “quality of urban infrastructure and services”: the roads, drains, traffic, transport, water supply, etc., in our urban areas. The second aspect of “quality of life” is the “quality of citizenship”: the role that urban residents play by participating in civic issues, deepening democracy and holding our public institutions accountable in delivering various goods and services. This second aspect includes very importantly the aspect of voluntarism. It runs on a technology-based project called iJanaagraha,¹¹⁸ which focusses on the using maps in analysing ward-level infrastructure and service data in Bangalore.¹¹⁹

Mapunity develops technology to tackle social problems and development challenges. Their GIS, MIS and mobile technologies are used mostly by government departments and civil society organisations, and in the R&D initiatives of commercial ventures. The company is incubated at International Institute of Information Technology, Bangalore and at the Indian Institute of Management, Bangalore. They have developed the traffic management system in many cities including the Bangalore Traffic Information Service (<http://www.btis.in>), and a portal for the municipal body elections in Bangalore. They work mostly on using technology to tackle social problems and development challenges in India. They work with GIS, MIS, and mobile technologies, mostly to be used by government agencies and civil society organizations.¹²⁰

The **Office of the Adviser to the Prime Minister** on Public Information Infrastructure and Innovations undertakes the task of reviewing, developing, utilising and scaling public information infrastructure in the country to help improve productivity, efficiency and quality of the systems and processes to deliver public services for citizen empowerment. The Office of Adviser will discuss, debate, analyse, articulate, and sensitise the need to innovate, at all levels and in all sectors in the country with a focus on inclusive growth, global competitiveness and prosperity, and create a Roadmap for a Decade of Innovation to meet the challenges of the 21st century.¹²¹

Praja Foundation is a non-partisan organization. It was started in Mumbai in 1997 by a group of eight individuals whose vision was to re-establish accountability and transparency in governance. These individuals were also concerned about a general lack of interest among the Citizens'

¹¹⁶<http://indiabiodiversity.org/about>

¹¹⁷http://indiabiodiversity.org/sites/default/files/lbp_geomatrix_paper.pdf

¹¹⁸<http://www.ijanaagraha.org>

¹¹⁹<http://www.janaagraha.org>

¹²⁰<http://www.mapunity.org>

¹²¹<http://www.iii.gov.in>

in the local government. Their initial efforts were concentrated on improving the capacity of the Brihanmumbai Municipal Corporation (BMC) to deliver better services and to be responsive to the people. Over time, Praja became a pressure group persistently advocating for a pro-citizen work ethic.¹²²

Praja.in is a community-driven loose-knit organisation. It describes itself as aiming “to be a bridge between those who serve us and those amongst us who care and want to participate. Towards that, it aims to establish an Internet driven community to help make the connection at local levels. It wants to be a networking platform for active and concerned citizens.”¹²³

Sadhana was conceptualised in 1992 by a young group of educationists and activists to explore alternative and creative approaches to educate, make education more child-centered and provide equal opportunities of education to all children regardless of caste, creed and colour. It is currently implementing a wide range of programmes in Medak district of Andhra Pradesh to mobilize communities around the issues of universalisation of education and children’s rights.¹²⁴

SmartVote.in works on issues of civic awareness and improvement. SmartVote.in is a citizen’s initiative to create a better civic life in their neighbourhoods through collective action. They helped propagate candidate information and conducted interviews with most of the candidates from Bangalore during the 2009 general elections. They focussed on a single constituency within Bangalore (Koramangala) during the municipal body elections in 2010.¹²⁵

Teachers of India (TOI), set up by the Azim Premji Foundation, an NGO working “in the area of elementary education to pilot and develop 'proofs of concept' that have a potential for systemic change in India's 1.3 million government-run schools.”¹²⁶ Teachers of India provides “a space for teachers to express their ideas and share their thoughts on any subject that touches their professional lives [and hosts] material created by and for teachers in English and vernacular languages”¹²⁷ including Hindi, Kannada, Tamil, Telugu, Oriya, Marathi, Gujarati, Malayalam, and Punjabi.

¹²²<http://www.praja.org>

¹²³<http://praja.in/about-praja>

¹²⁴<http://www.sadhanango.org>

¹²⁵<http://www.smartvote.in>

¹²⁶<http://www.azimpremjifoundation.org/vision.html>

¹²⁷<http://www.teachersofindia.org/>

Annexure II: List of People Interviewed and Interacted With

Office of the Adviser to the PM, Public Information Infrastructure and Innovations

Sam Pitroda

sam.pitroda@c-sam.com

AGRISNET Tamil Nadu

Thiru P. Venkatachalapathy Agricultural Officer (GOI Schemes)-I.

diragri@tn.nic.in

Akshara Foundation

Gautam John, New Projects Manager

gautam@akshara.org.in

Alok Singh, Technology Lead

alok@klp.org.in

Shivangi Desai

shivangi@klp.org.in

Asia and Pacific Centre for Transfer of Technology

N. Srinivasan, Director, Innovation Management

srini@apctt.org

Association for Democratic Reforms

Anil Bairwal, Executive Director

anil@adrindia.org

BMTC Bus Planner

Mithila Jha Urban Planner

09739097539

Busroutes.in / Geohackers

Arun Ganesh, Developer

arungraphy@gmail.com

Central Information Commission

Shailesh Gandhi, Information Commissioner

shaileshgandhi@nic.in

Centre for Budget and Governance Accountability (CBGA)

Subrat Das, Executive Director

subrat2005@gmail.com

Centre for Good Governance (CGG)

T. Vijay Karan Reddy, Senior Project Manager, e-Governance

vijaykaran@cgg.gov.in

N. Damodar Abbai, Senior Project Manager

Accountability Initiative, Centre for Policy Research (CPR)

Mandakini Devasher, Research Associate

mdevasher@accountabilityindia.org

Sruti Bandyopadhyay, Research Associate
sbandyopadhyay@accountabilityindia.org

Shibani Ghosh, Part-time Research Associate
shibanighosh83@gmail.com

Citizen Matters

Subramaniam Vincent, Co- Founder, Reporter
edit@citizenmatters.in
Navya P K, Reporter
navya@oorvani.in

Commonwealth Human Rights Initiative

Venkatesh Nayak, Coordinator, Access to Information Programme
venkatesh@humanrightsinitiative.org

Copyright Office

G.R. Raghavender, Deputy Secretary (Ministry of HRD) & Registrar of Copyrights
raghavender.edu@nic.in

Data Security Council of India

Vinayak Godse, Director of Data Protection
vinayak.godse@dsci.in

Department of Information Technology, Ministry of Communications and Information Technology

N. Ravishankar, Joint Secretary
nravishanker@mit.gov.in

Shankar Aggarwal, Joint Secretary
sagg-js@mit.gov.in

Dr. Govind, Director, E-Infrastructure & Internet Governance Division
drgovind@nic.in

Digital Green

Rikin Gandhi
rikin@digitalgreen.org

Ekgaon

K Sasikumar
sasi@ekgaon.com

Election Commission of India

Salek Chand, Library and Information Officer
feedback@eci.gov.in

Google India

Raman Chima, Policy Analyst
ramanjit@google.com

Governance Now

B.V. Rao, Editor
bvrao@governancenow.com

Samir Sachdeva, Assistant Editor
samir@governancenow.com

IndiaGoverns Research Institute

Veena Ramanna, Executive Director
veena@indiagoverns.org

India Governance Knowledge Centre

Sapna Kedia, Research Assistant
Aryamala Prasad, Research Assistant
Attrika Hazarika, Research Assistant

India Water Portal

Deepak Menon, Manager
deepak@indiawaterportal.org

Indian Kanoon

Sushant Sinha
sushant@indiankanoon.com

IT For Change

Parminder Jeet Singh
parminder@itforchange.net

Janaagraha

Thejesh GN, Technology Manager
thej@janaagraha.org

Kerala IT Mission

Sabaraish
sabarish@keralaitmission.org

Kiirti

Sudha Nair, Project Manager
sudha@emoksha.org

Empowering India

Barun Mitra, Managing Trustee
barunmitra@gmail.com

Mahiti Infotech

Vijay Rasquinha, Chief Technology Officer
vijay@mahiti.org

Chethan Elvis Das, Marketing Director
chethan@mahiti.org

Mahita Adhikar Manch

Bhaskar Prabhu, Director
prabhuisbhaskar@gmail.com

Maharashtra Department of Agriculture

Balasaheb Thorat, Commissioner
agriipa.mah@nic.in

Maharashtra Department of Environment

ENVIS Coordinator
mah@envis.nic.in

Mapunity

Sridhar Raman, Software Engineer
sridhar@mapunity.in

S. Raghavan Kandala, Software Developer
raghavan.kandala@mapunity.in

Ministry of Agriculture, Government of India

Utpaul Ghosh, Economics and Statistics Adviser
eco-adv.moca@nic.in

The National Centre For Advocacy Studies (NCAS)

Shirish N. Kavadi, Research Director
shirish.kavadi@gmail.com

National Informatics Centre (NIC)

Dr. B. K. Gairola, Director General
dg@nic.in

National Institute for Smart Government (NISG)

T. Vijay Saradhi, Vice President
vijays@nisg.org

Satyajit Suri, General Manager
satyajit.suri@nisg.org

Amit Chakrabarty, Projects Manager
amit.chakravarty@nisg.org

National Network For India (NNFI)

Anant Trivedi
trivedianant@gmail.com
Promod Chawla
promodchawla@gmail.com

National Spatial Data Infrastructure (NSDI)

Maj. Gen. Dr. R. Siva Kumar, CEO, NSDI & Head, Natural Resources Data Management System
siva_k@nic.in

OpenCivic

Akshay Surve, Developer
akshay.surve@gmail.com

PRS Legislative Research

Chakshu Roy, Analyst
chakshu@prsindia.org

Praja Foundation

Nitai Mehta, Managing Trustee
nitai@aditya.net

Reserve Bank of India

Dr. A.M.Pedgaonkar, Principal Adviser, Department of Statistics and Information Management
C-8/9, Bandra-Kurla Complex
Post Box No.8128, Bandra (East)
Mumbai-400 051.
ampedgaonkar@rbi.org.in

Sadhana

Murali Mohan, Executive Director
sadhanamurali@yahoo.co.in

SmartVote.in

P.G. Bhat, Volunteer
info@smartvote.in

Tamil Nadu IT Department Secretary

Santosh Babu, Secretary
secyit.tn@nic.in

Thiruvananthapuram Municipal Corporation

Biju K IAS Secretary
tvpmcorp@gmail.com

Transparency International Tamil Nadu

L.M. Menezes I.A.S. (Retd.)
tiitn77@gmail.com

Transparent Chennai

Nithya V Raman
nithya.raman@ifmr.ac.in

Yashwantrao Chavan Academy of Development Administration (YASHADA)

Pralhad Kachare, Director, Center for Public Policy
pkachare@gmail.com