Opening Government Data through Mediation: Exploring the Roles, Practices and Strategies of Data Intermediary Organisations in India

Sumandro Chattapadhyay
http://ajantriks.net/
mail@ajantriks.net

Project Report
http://ajantriks.github.io/oddc/

Contents

1. Introduction

The National Data Sharing and Accessibility Policy[1] (henceforth, NDSAP) was approved by the Government of India in early 2012 to encourage and govern publication of government created and owned datasets, from various agencies of the central government of India, in open digital formats and through a single national data portal. The Policy extended the mandate of the Right to Information Act of 2005[2] by establishing the principles and procedures of proactive disclosure of government data in India. The national data portal of India, or the Open Government Data Platform of India <http://data.gov.in>, was launched later in 2012 to start collecting and publishing open government data. The NDSAP Project Management Unit (henceforth, NDSAP-PMU), within National Informatics Centre, was entrusted with development, deployment, and management of the Platform and popularising publication and usage of open government data. Over the last years the NDSAP-PMU has organised several outreach events, workshops and competitions, often in collaboration with different industrial and academic bodies, towards these goals.

The open government data agenda and its implementation in India, however, remains still too young to make possible a study of its outcomes and impacts. This study, hence, explores not the outcomes of the NDSAP or the Open Government Data Platform of India as such, but the existing practices of accessing and using government data in India to understand what challenges this Policy and its implementations should respond to, and what available opportunities can be mobilised towards an effective open data agenda.

---

1   Department of Science and Technology, Government of India. 2012. National Data Sharing and Accessibility Policy. *The Gazette of India*. March 17. 74-99. http://data.gov.in/sites/default/files/NDSAP.pdf
2   Ministry of Law and Justice, Government of India. 2005. *Right to Information Act, 2005*. Act No. 22 of 2005. http://righttoinformation.gov.in/rti-act.pdf

Beginning in May 2013 as part of the 'Exploring the Emerging Impacts of Open Data in Developing Countries' (ODDC) research network managed by the World Wide Web Foundation and supported by the International Development Research Centre, Canada, this study was initially conceptualised by the author and Zainab Bawa of HasGeek Media LLP as a comparative study of organisations working with urban data in India. The proposed project originally planned to locate these policy documents, especially NDSAP, in the context of their actual implications and uses for non-government data practitioners. We wanted to focus on non-government advocacy and research organisations working in the area of urban development across five cities in India – Ahmedabad, Bengaluru, Delhi, Mumbai and Pune – and to map their practices of collecting, accessing, sharing and using government and self-generated urban data, such as those related to property records, geo-spatial data, sanitation and public health, elected representatives etc. The two key questions for the study were: (1) how RTI and NDSAP have transformed approaches and processes of data practices, and (2) can the grounded experience of advocacy and research organisations inform shaping of a more effective and open data policy for India?

However, two principal factors compelled us to rethink the approach of the study. Firstly, our initial discussions with policy researchers, RTI activists, and open data activists revealed possibility of curious disjunction between the open data and the Right to Information movements in India. This is especially interesting in our specific policy context where the NDSAP document gets developed as an extension of the RTI act. Given this early finding, we were keen to revisit our first research question (mentioned above) and to categorically explore the experiences of accessing government data by various organisations in India, to specifically highlight out the relative uses of (or not) of the RTI act and the NDSAP for this purpose. Secondly, the subject area of urban development

was originally chosen to systematically focus the scope of the study, as well as to chose a specific governance context to gather more precise insights. Our initial discussions with policy researchers, open data activists and RTI activists indicated that by limiting the study to a certain subject area, we might be losing out on interacting with organisations that are key actors in the domain of accessing government information and data in India.

For example, Arghyam <http://www.arghyam.org/> works on issues of water and sanitation and is at the forefront of accessing, using and opening up water related government data for further uses; Karnataka Learning Partnership <http://klp.org.in/> focuses on primary education challenges in Karnataka (a southern state of India) and gathers substantial primary data to analyse performance in primary schools across the state; and Association for Democratic Reforms <http://adrindia.org/>is singularly responsible for ensuring public sharing of affidavits (including data on social, economic, and criminal records) submitted by political candidates before participating in elections. We were of the opinion that these non-government organisations, of diverse sectoral foci, must be engaged with to have a more robust understanding of how government data and information is being accessed, used and opened up in India.

The reformulated study, hence, decided to discard the thematic focus in selection of organisations to be studied, and to look into organisations that access, use, and share national level (since NDSAP is a national level, and not a state level policy) government data. It focused on exploring the actual practices around government data by various (non-governmental) "data intermediary organisations" on one hand, and implementation challenges faced by managers of the Open Government Data Platform of India on the other, so as to identify possible areas of policy modification, capacity building, community organisation, and alignment of efforts.

The study has the following objectives:

- To undertake a provisional mapping of government data related activities across different sectors to understand the nature of the "open data community" in India,

- To enrich government data/information policy discussion in India by gathering evidence and experience of (non-governmental) data intermediaries regarding their actual practices of accessing and sharing government data, and their utilisation of the provisions of NDSAP and RTI act, and

- To critically reflect on the nature of open data practices in India.

2. Methodology

2.1. Why Data Intermediary Organisations?

The idea of looking for the "data intermediary organisations" in India began to take shape during the reformulation of the study plan after the first Network Meeting in London and through early discussions with several analysts and members of open data and Right to Information groups in India. Further, conversation with Michael Gurstein, the mentor of the study, led to a conceptualisation of the category of "data intermediary organisation" inspired by the idea of (inter)mediation both in the context of local development – where the intermediary organisation would help local communities to engage and transact with various service delivery agencies (public and private) more effectively – and also in the context of information – where the intermediary organisation would focus on improving the access to and use of information for its intended audiences. In the Indian context it became clear early on that national- or state-level government

data, once published by the agency concerned, most commonly travel through a set of research and advocacy organisations before it is consumed by public, either in print or in digital formats. This study focuses on a selected subset of such organisations.

In international discussions of the open data ecosystem, and surveys of the typology of intermediaries populating that ecosystem, often tend to identify a single set of organisations (mostly government agencies) that undertakes the supply of open data, while other types of organisations perform various kinds of value-addition to the data and deliver data-based products and services.[3] This imagination neither reflects the Indian situation, where several non-governmental organisations play a key role not only in self-collection of data but also in converting closed government data or openly available information into usable data, nor does it set up an aspirational standard to indicate how the open data ecosystem should evolve. Rufus Pollock of Open Knowledge Foundation offers a sharp criticism of such "one way street" imaginations of the open data ecosystem.[4] To illustrate Pollock's argument, we may think of the adjective 'open' as applying to both "data" and "ecosystem." In other words, open data ecosystem needs to be conceptualised (and realised) as a network of creators and users of open data, where there is no unidirectional flow of open data.

In an open data ecosystem, the data is shared by its creators and also by its users. The government agencies, in an open data ecosystem, may not only share open data but also get back in return value-added analysis, insights, and datasets from a wide-range of users of data. A key feature of an open ecosystem is that it allows for resources to be

---

3   Magalhaes, Gustavo, Catarina Roseira, and Sharon Strover. 2013. Open Government Data Intermediaries: A
     Terminology Framework. *In Proceedings of the 7th International Conference on Theory and Practice of Electronic
     Governance (ICEGOV '13)*, Tomasz Janowski, Jeanne Holm, and Elsa Estevez (Eds.). New York, NY: ACM. DOI=
     9781450324564/00/0010.
     Deloitte Analytics. 2012. *Open Growth: Stimulating Demand for Open Data in the UK. Briefing Note*.
     http://www.deloitte.com/assets/Dcom-UnitedKingdom/Local%20Assets/Documents/Market%20insights/Deloitte
     %20Analytics/uk-da-open-growth.pdf
4   Pollock, Rufus. 2011. Building the (Open) Data Ecosystem. Open Knowledge Foundation Blog. March 31.
     http://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/

available in various forms so that all the constituting members of the system may access and consume it in the form appropriate for the member concerned. These users often require data at different scales of granularity and expanse (spatial and/ temporal) that enable situational use driven by the context of final users. Availability of such data requires all or at least a diverse range of actors in the ecosystem that publish primary and/or value-added data.

In their foundational work, John Hagel III and Jeffrey Rayport predicted the rise of "infomediary" organisations, which will offer personal and agency-wide management as a service, and enable individual producers of data to minutely control the access to their data by various agents (that may provide data-based products and services against that).[5] On the other hand, the economics of information literature has discussed in detail the characteristics and socially optimal form of "information intermediary" organisations.[6] Such an organisation is defined by its information processing activities (can be both for-profit or otherwise) being determined by information needs of its clients. The idea of an "information intermediary" organisation has been also discussed in a developmental context. There the focus has obviously been not only on building platforms to store and share information but also on active explorations in understanding how communities (and individuals within them) process information, and thus how information should be shared for greatest possible impact.[7]

---

5   Hagel III, John and Jeffrey F. Rayport. 1997. The Coming Battle for Customer Information. *Harvard Business Review*. January-February. Pp. 6-11.
6   Rose, F. 1999. *The Economics, Concept, and Design of Information Intermediaries: A Theoretic Approach*. Heidelberg, Germany: Physica-Verlag.
     Womack, Ryan. 2002. Information Intermediaries and Optimal Information Distribution. *Library and Information Science Research*. 24. Pp. 129-155.
7   Fisher, Catherine. 2010. *Five Characteristics of Effective Information Intermediary Organisations and How to Ensure You Have Them*. Institute of Development Studies (IDS) Knowledge Services. http://www.ids.ac.uk/files/dmfile/5_characteristicsofintermediaries_EWUpdate.pdf

The concept of "data intermediary organisation" employed in this study is closest to this developmental understanding of "information intermediary" organisations. The replacement of "information" by "data" indicates the emphasis put on mediation of access to open data itself, and not only of information to be directly consumer by intended audiences. Further, it is important that the provision of open data is not only driven by demands of pre-identified client groups. These data intermediaries are, thus, expected to enhance the quality and amplify the circulation of data opened up by the government agencies, through acts of sanitising, organising, compiling, formatting, and documenting available open government data. These organisations may also additionally function as repositories of open data sourced from non-government actors. Thus, within the overall ecosystem, open data intermediaries will create focused – either regional or sectoral – loops of data flow and value-addition and augment the ecosystem as a whole.

2.2. Data Intermediary Organisations

In this study, a "data intermediary organisation" is conceptualised as an organisation that shares data for its access, consumption and re-usage (including re-sharing) by other organisations and individuals. Three further clarifications are needed here:

- Sharing of data by such organisations can either be done on a commercial or a non-commercial basis,

- Shared data can either be sourced from an external creator and publisher of data (either government or private agencies), or be created by the data intermediary organisation itself, and

- The data intermediary organisation may or may not add (or reduce) value of the data before sharing it further, that is it may or may not modify the data received

by them (by cleaning up, re-organising, re-formatting, aggregating, etc.) before sharing it.

This study recognises that given the lack of (hierarchical) depth of the (government and non-government) data access and usage ecology in India, often the same organisation (especially the smaller organisations) are compelled to undertake multiple data-related functions internally. Instead of categorising organisations that we will study as "data creating and publishing organisation," "data intermediary organisation" and so on, the study identifies the functions performed by the organisation concerned. It surveys several (but surely not all) organisations that mediate access to and use of government data in India. Many of these organisations are expected to perform multiple data-related functions such as creating data, directly using data to inform organisational activities, sharing the data with other organisations and citizens in general, training other organisations and individuals to use (collect, analyse, etc.) data, etc. However, an organisation can be understood as a "data intermediary organisation," in the context of this study, if it considers (and its activities reveal) "data intermediation" as one of its key functions. Since a direct understanding of the primary function(s) of an organisation is often difficult to obtain, this problem is approached through various questions (in the conversation with the organisation concerned) so as to triangulate towards the identification of its various functions.

But most importantly, the objective of this study is not to evaluate the surveyed organisations and understand if they are "data intermediary organisations" or not, but to understand their experiences in accessing and sharing (so as to mediate access) government data, and what opportunities and challenges they face in functioning as "data intermediary organisations."

2.3. Identifying Organisations

As discussed above, this study has not been limited only to organisations that purely function as "data intermediary organisations" but interacted with a wider range of organisations that perform that task of mediating access to data and information and enabling its usage by other organisations. Many of the organisations (interacted with in the study) perform multiple data-related functions such as creating data, using data to inform intra-organisational activities, sharing data with other organisations and citizens in general, training other organisations and individuals to use (collect, analyse, etc.) data, etc. A few of the organisations considered in the study do not share data with other organisations at the present, but have the potential – in terms of willingness and capacity – to do so. As the study was not interested in assessing and identifying the organisations in India that function as "data intermediary organisations" but in understanding the existing range of practices of accessing, using, and re-sharing national-scale government data in India, it was crucial to consider a selected set of organisations that differ in their sectoral focus, strategy and theory of change, technological capacities and geographic location. That being said, these organisations were needed to be involved in working with national-scale government data sets and in publishing them for public usage – with or without restrictions – not only as information but also as data.

To decide upon the initial set of organisations to begin the study with – as it was expected that snowball sampling will be most useful to identify the entire set of organisations to be considered in this study – an Internet-based survey was launched on June 25, 2013[8]. In this survey, representatives of organisations that work with significant amount of government data were invited to submit the following information:

---

8   See: http://ajantriks.github.io/oddc/posts/2013.06.25_survey.html

- Details about the respondent and the organisation concerned,

- Whether the organisation identify itself as functioning as "data/information intermediary,"

- Details about how the organisation works with data and/or information,

- Whether the organisation would like to be interviewed for this study, and

- Whether there are other "data intermediary organisations" known by the respondent who should be considered in this study.

Link to this survey was shared through various mailing lists and professional networks in India that are related to questions of data, open data, and development. However, it attracted very few responses and not the most useful ones. There were five organisations that completed this survey: Enthought Inc., Fields of View, Indian Institute of Human Settlements, Karnataka Learning Partnership, and Niti Digital. While Enthought Inc. and Niti Digital are commercial organisations that substantially work with data to produce analytics, visualisations, reports, and services, they do not publish the same data in re-usable formats. Fields of View and Indian Institute of Human Settlements are research and advocacy organisations that are also involved in strong data-driven research activities but neither do they undertake significant re-sharing of the data they work with. Interestingly, it is Gautam John of Karnataka Learning Partnership, which is one of the most important "data intermediary organisations" in India, who also suggested names of other organisations that this study can potentially survey. The experience of this survey suggested that the most effective way to proceed with this study would be to start speaking to organisations, such as Karnataka Learning Partnership, that are most well-known within the open data circuit in India and follow the references to other organisations offered by them.

Two more decisions were taken at this point. Firstly, the initial Internet-based survey and online/offline searches did not reveal any commercial organisation in India that accesses, uses, and re-shares (as opposed to re-selling) government data. Hence the focus of the study was narrowed down to non-governmental research and advocacy organisations. Secondly, the five national information/knowledge portals of India – India Biodiversity Portal run by Strand Life Sciences <http://indiabiodiversity.org/>, India Energy Portal run by The Energy and Resources Institute <http://indiaenergyportal.org/>, India Environmental Portal run by Centre for Science and Environment <http://indiaenvironmentportal.org.in/>, India Urban Portal run by National Institute of Urban Affairs <http://indiaurbanportal.in/>, and India Water Portal run by Arghyam <http://www.indiawaterportal.org/> – were automatically selected for consideration in this study, since they already play a critical role in aggregating, archiving and sharing large sector-specific bodies of information and have a great potential to also host and share government and non-government data related to the sectoral issues (if they are not doing it already).

The methodological decisions mentioned above clearly creates a possibility of introducing subjective bias in the selection of organisations to be studied. To counterweight this, all the organisations interacted with during the study were asked to suggest names of other organisations that should also be addressed by the study. Answers to this question largely generated references back to organisations that have already been (or planned to be) interacted with, and thus validated the inclusion of the organisations concerned. The study comprised of engagements with fourteen organisations, located across Bengaluru (three organisations), Chennai (one organisation), Delhi (nine organisations), and Hyderabad (one organisation). There is an explicit dominance of organisations from Delhi here, primarily due to the concentration of research and advocacy organisations in the capital. Survey of five organisations from three cities in southern part of India partially

balance the overall bias towards organisations from Delhi, that is north India. In terms of thematic focus, the surveyed organisations work across the following sectors: budget and governance expenditure (two organisations), education (two organisations), electoral and parliamentary transparency (two organisations), environment (five organisations), and urban development (three organisations). In terms of sectoral involvements, the set of organisations included in this study are thus quite diverse and help government data insights to be brough together from across these sectors. For lack of time, a small number of organisations were initially considered but finally not included in the study, such as Bangalore Urban Metabolism Project, Digital Green, and Praja Foundation.

2.4. List of Organisations Surveyed

The organisations surveyed in this study are:

- Accountability Initiative, Centre for Policy Research, Delhi

- Akvo Foundation, Delhi

- ASER Centre, Delhi

- Association for Democratic Reforms, Delhi

- Centre for Budget and Governance Accountability, Delhi

- Hyderabad Urban Lab, Hyderabad

- Karnataka Learning Partnership, Bengaluru

- India Biodiversity Portal, Strand Life Sciences, Bangalore

- India Energy Portal, The Energy and Resources Institute, Delhi

- India Environment Portal, Centre for Science and Environment, Delhi

- India Urban Portal, National Institute of Urban Affairs, Delhi

- India Water Portal, Arghyam, Bengaluru

- PRS Legislative Research, Delhi

- Transparent Chennai, Chennai

Further, the following individuals were interviewed as part of the background research for this study:

- Ms. Alka Mishra and Mr. Durga Prasad Misra of National Informatics Centre,

- Ms. Nisha Thompson of DataMeet, and

- Mr. Pranesh Prakash of The Centre for Internet and Society.

2.5. Survey Questionnaire

The questionnaire used in the interviews was developed first during May-June 2013. The draft questionnaire was presented to the ODDC research network in general and the participants of the Asian Regional Meeting (held in Delhi in July 2013) in particular. Various critical responses were received and incorporated to prepare the final questionnaire. The questionnaire is divided in six sections:

- Organisation: About the organisation and its involvement with open data

- Accessing Data: About accessing data from government and non-government origins

- Using Data: About usage of data internally and in relationship with other organisations

- Collecting Data: About data created by the organisation

- Sharing Data: About sharing of data by the organisation

- Data Ecology: About the organisation's linkages with other organisations in the matter of accessing, using, and sharing data

Please refer to **Section 6.3.** for the full questionnaire.

3. Open Government Data Policy and Platform in India

3.1. National Data Sharing and Accessibility Policy

Since late 1990s, the Indian state has rapidly, though unevenly, moved towards extensive adoption of information and communication technologies to raise the quality and ease of state-citizen interactions. The "Report of the Working Group on Convergence and E-Governance for the Tenth Five Year Plan (2002-07)" gave an early call to arms for electronic governance initiatives that "involves transformation from being a passive information and service provider to active citizen involvement."[9] While the Right to Information Act of 2005 created a reactive disclosure instrument for the citizens to realise their right to gather information about governmental activities, processes and their status, a crucial gap remained in terms of the citizen's ability to access government data in large scales and easily usable formats. This lacuna was addressed by the National Data Sharing and Accessibility Policy (NDSAP) prepared by the Department of Science and Technology in 2012. This is the first policy in India that categorically deals with access to public data in machine-readable formats.

---

9   Planning Commission, Government of India. 2001. *Report of the Working Group on Convergence and E-Governance for the Tenth Five Year Plan (2002-07)*. New Delhi: Government of India. P. 34.

During the 2010-11 India visit of Barack Obama, President of the United States of America, several collaborations between the Indian and the US governments were launched. One of such initiatives was the Open Government Platform – an open source data and content management system that can be customised easily to develop open data portals for various types of agencies, including national government. National Informatics Centre, Ministry of Communications and Information Technology, Government of India (henceforth, NIC) and Office of Citizen Services and Innovative Technologies, General Services Administration, Government of USA, undertook a collaborative software development project to build the Open Government Platform in early 2011. It is to be noted that this software development project and drafting of the NDSAP proceeded side by side, and the requirements of one informed the contents of the other. As the text of the NDSAP document makes explicit, the dominant focus of the Policy is towards describing the desired functioning of the Open Government Data Platform of India – the URL of the same was specified in the Policy itself – as opposed to describing a government-wide re-engineering of data collection, management and publication practices towards opening up government data.

The first draft version of the NDSAP document was published first on May 2011 along with a call for public responses. Although this may have attracted various feedbacks from different stakeholders, only the one submitted by The Centre for Internet and Society, Bengaluru, is publicly accessible (from its website).[10] None of the suggestions from this submission, however, were reflected in the final version of the policy published in The Gazette of India on March 17, 2012. The policy declared that data produced, collected and collated by the government agencies using public funds should be made publicly available in an organised, well-documented and timely manner, so as to enable the use of such data to produce socio-democratic as well as economic value.

---

10 Prakash, Pranesh. 2011. Comments on the Draft National Data Sharing and Accessibility Policy. The Centre for Internet and Society. June 08. http://cis-india.org/openness/blog/draft-ndsap-comments

The policy covers all ministries, departments, subordinate offices, organisations and autonomous bodies of the central government, and mandates that all "shareable" "non-sensitive" data should be published through a common government data portal deployed and managed by NIC. Here "shareable" refers to data declared to be such by the government agency that created it, and "non-sensitive" refers to data sharing of which is not prohibited by any central government acts.

NDSAP further states that the published data must be available in both human-readable and machine-readable formats, adopt file and metadata standards as specified by NIC, and should be updated regularly. Though this policy was directly targeted at central government agencies, it was expected that equivalent state-level policies will be passed subsequently, which would facilitate open sharing of data across various tiers of the government. While the policy upholds the principle of "openness," in actuality it only enforces use of open standards for publication of data and not the simultaneous adoption of open licenses. Thus, strictly speaking, NDSAP cannot be called an *open* government data policy. Instead, it must be seen as the first step taken by the Indian government to consider born digital public data as a special category of government data, thus requiring a specific manner of its publication, and acceptance of the unique value of machine-readable datasets for effective citizenship.

Moreover, an effective open government data policy must not only ensure re-usability of public sector information and open data standards to enable such reusability, but must also be integrated with various administrative programmes driven by and generating significant amounts of data about governmental processes. It needs to be formulated within specific national policy context, considering the varying administrative cultures and governance frameworks to collect, archive, process and make accessible digital traces

of citizenship, statecraft and state-citizen interactions. NDSAP falls short of substantially opening up government data in this respect too, as it continues to lack the element of interdepartmental and cross-policy referencing to build a more robust and coherent programme of open government data and access.

3.2. Open Government Data Platform of India

Following the directive of NDSAP, NIC created the NDSAP-PMU (NDSAP Project Management Unit) in 2012 to build the Open Government Data Platform of India <http://data.gov.in/>. The first version of the Platform was launched in May 2012. It was powered by a thoroughly customised version of the above mentioned Open Government Platform software. Since then the NDSAP-PMU has interpreted its initial mandate of developing and managing the data portal to include various other critical roles such as enforcing adoption of "open standards" for published datasets, evangelising proactive sharing of government data across agencies through detailed consultation meetings, organising community outreach programmes to induce increased usage of the datasets available from the portal, etc. Informed by its consultations with various government agencies, non-government organisations and citizen groups, it prepared an implementation guidelines document that has already been through a few iterations.[11] This document details out the stages of the data contribution process, including the role and responsibilities of the Data Controllers, relevant metadata and file standards, and management of datasets after they have been published in the portal.

---

11 Open Government Data Division, National Informatics Centre, Ministry of Communications and Information Technology, Government of India. 2014. *Implementation Guidelines for National Data Sharing and Accessibility Policy (NDSAP) - Version 2.2*. February.
http://data.gov.in/sites/default/files/NDSAP_Implementation_Guidelines_2.2.pdf

The most recent version of the Platform was launched on February 18, 2014. The website is now powered by an upgraded Open Government Platform built on Drupal 7[12] that implements API based public access to the datasets hosted in the platform. The Platform attempts to provide an unified catalog of data sets submitted to the Platform by various central government agencies. In most cases, the submitted data is hosted at and made available from the Platform itself, while in a few rare cases the Platform provides link to data actually hosted in the server of the data publishing agency. All central government agencies – Ministries and organisations within it – have chosen a Data Controller to drive and coordinate the data publication activities of the agency concerned. Numerous Data Contributors work within each such agency, led by the Data Controller, to submit data sets to the Platform through a web-based submission and scrutiny process. NDSAP-PMU is in the process of deliberating with various state governments in India about the possibility of launching state-level data portals powered by the Open Government Platform software, deployed in the Platform-as-a-Service model.

This software allows users, both governmental and non-governmental, to browse the data catalog, view the metadata associated with each data set, comment on and rank various aspects of the data set, create basic visualisations by choosing variables from the dataset, download available data sets and submit request for those that are not available yet. Further, it has features such as for exporting of published data sets in various open formats (such as, CSV and XML) irrespective of the format of the original data sets. After initially using the Recline.js library <http://okfnlabs.org/recline/>, developed by Open Knowledge Foundation, for within-Platform visualisation of hosted data sets, it has recently launched a self-built visualisation engine <http://data.gov.in/visualize3/>. This engine allows visualisation of hosted data sets as linkable and embeddable charts, and can also be used to create menu-driven visualisations of data sets supplied by the user.

---

12 See: https://github.com/opengovplatform/OGPL-D7.1_alpha

4. Key Themes

4.1. Open Data Community in India

One of the earliest events that brought together people talking about, or perhaps wondering about, open data in India was the barcamp on "Technology, Transparency and Accountability" organised by Accountability Initiative <http://accountabilityindia.org/>, Centre for Policy Research, on June 05, 2011 in Gurgaon.[13] The event was critical in bringing together more experienced information activists, especially those connected with the Right to Information movement, and relatively newer participants in transparency and accountability discussions in India, especially those who brought in a substantially technology-oriented approach. Several participants of this barcamp have since played important roles in the making of the open data agenda and practices in India. However, the interactions among the Right to Information activists and groups in India with those working with open data are yet to be become significant.

One of the key questions for this study was to understand the shape and the dynamics of "open data community" in India. Although majority of respondents recognised that a particular network – DataMeet <http://datameet.org/> – or specific persons associated with the network as the people who are driving the open data conversations in India, they also noted that their organisations consider themselves to be on the "sidelines" of this "open data community." Two reasons were commonly offered for such an identification:

---

13 Accountability Initiative. 2011. Technology, Accountability & Transparency Camp.
    http://www.accountabilityindia.in/events/2239-technology-accountability-transparency-camp

- The kinds of government data that the organisation works with are already publicly available, though not necessarily in open format, and hence the organisation do not feel the need be involved in the "open data community" in a rigorous way, and

- The kinds of data that the organisation work with are not collected by the government at all, and hence it does not spend any organisational resources to push the government towards proactive disclosure of data.

The above reasonings, however, do not imply that these organisations are not interested in the open data agenda. On the contrary, all of these organisations that considered themselves to be on the fringes of the "open data community" explicitly stated that they consider open government data to be a necessary (although not sufficient) resource for ensuring greater transparency in governmental activities. Given limited organisational resources, however, most of the organisations surveyed here do no undertake more involved roles within the "open data community," as they mostly work with data that is already publicly available, or with data not collected by the government at all. Further, it is of great interest that all organisations surveyed in this study considered themselves either as part of the "open data community" in India, or *natural (or "by default") allies* of the same. This also reflects a sense of ambiguity regarding the very conditions of *belonging* to this "open data community." Anticipating such a situation,this study did not employ any fixed determinants of *belonging* but depended on self-identification by the organisations. Anant Maringanti of Hyderabad Urban Lab <http://hydlab.in/> noted that, although, the commercial re-users of government data and the transparency and accountability activism interests have converged in North American and European countries to shape the "open data agenda," which is yet to be the reality in India. Neither the commercial re-users of government data have organised itself to articulate a clear

demand for open data practices by the government, nor have large-scale data-driven monitoring (by non-government agencies) of transparency and accountability of government activities become commonplace. It is this condition that creates the lack of clear understanding about what it means to belong to "open data community" in India.

As Yamini Aiyar of Accountability Initiative explained, the "open data culture" in India is not only in its infancy but is mostly limited to a group of people who speak a certain vocabulary. This concern about the specific vocabulary, and technical knowledge and skills that go with it, of the people talking about open data has been flagged by several people interviewed in this study. This concern is heightened by the fact that though the open data agenda speaks a language that overlaps with related discourses of transparency and accountability on one hand, and evidence-driven developmental interventions on the other, it also deploys a precise technical understanding of digital data and its specific forms that enable easy programmatic manipulations. Gautam John of Karnataka Learning Partnership (who participated in the above mentioned barcamp) clarified that along with continuing focus on accountability and transparency issues, evidence-driven development practices are becoming more common in India, often driven by donor conditionalities and reporting requirements.

This is gradually generating a general culture of data usage within Non-Government Organisation and Civil Society Organisation spheres. It, however, may not always imply that these organisation see the data produced by them as means to (several) ends, and not only as an end product to be shared with the donor or the public in general. Furthermore, as Suman Bhattacharjea of ASER Centre <http://www.asercentre.org/> described, data collected by NGOs and CSOs for sharing with donors are often driven by the requirement for internationally comparable data sets, which are also sometimes

produced by government agencies monitoring delivery of various services. This effectively

leads to under-prioritisation of collection of data that is crucial for data/evidence-based

decision making and planning at the local and regional level. Such insights were shared

during the interviews often to illustrate the emerging data/evidence-based transparency

research and advocacy efforts in India, and the discontents thereof. These efforts create

the practical context of the "open data community" in India. The "community" struggles

on one hand to address the demands for data and related skills coming from such efforts,

and to envision a linked but autonomous agenda for open data in India.

Nisha Thompson, one of the earliest members and currently the Director of DataMeet,

narrated how the initial meetings of DataMeet, and the earlier conversations that led to

DataMeet, were largely driven by individuals across organisations and with an

understanding that (a) an effective engagement (and utilisation) of the issue of data, in

itself, requires a lot of time, efforts and support that most organisations are not in a

position to offer, and (b) the issue of transparency, understood in whichever way, is surely

related to that of (government) data but there is no clear notion of what that relationship

is. DataMeet, to the most extent, has been interested in addressing these concerns:

- Providing a space for sharing of knowledge and practices of working with, mostly
  government produced, data in India among members from various (commercial
  and non-commercial) organisations, so as to create an extra-institutional network
  (from the perspective of the organisations to which the DataMeet members belong)
  that supports data-related activities within the network, and

- Exploring and pushing the connections between transparency, accountability, open
  data, data journalism and data-driven public discussions in India.

Such a mode of functioning of DataMeet, the key open data community group in India,

also implies that open data sometimes remains an agenda driven by individual members

of an organisations – the members who are also associated with DataMeet – and not an organisation-wide agenda. This of course allows the organisations concerned to share learnings through the DataMeet network, but the organisation as an entirety does not become an actor in the open data community in India. Conversely, the open data agenda in India has been limited by the expanse of the DataMeet network.

Suman Bhattacharjea of ASER Centre argued that many organisations that work with a large amount of government data are yet to start talking about open data as such. This, however, is not only due to the incomplete popularisation of the open data agenda but also due to longer practices of treating data as an exclusive resource to harvest insights from. Further, P.K. Bhattacharya of The Energy and Resources Institute pointed out that the lack of government-wide streamlined processes for publication of government data either for public usage or as commercial products have also led to a prominent government data re-selling industry in India. Often these data re-selling companies publish data sets that appear to have been collected by government agencies but do not come with a clear attributing notice and documentation, since the companies access these data sets through informal routes. Such practices of liberating, re-organising, and formalising government data sets accessed through informal means or in closed formats are not only limited to the data re-selling firms, but are also significantly undertaken by various research and advocacy organisations. The lack of publication of government data in open, well-documented, and regular manner not only creates the space for a data re-selling industry, but also a "closed data culture" among non-commercial re-users of data. Several research and advocacy organisations surveyed in this study have mentioned that since they undertake substantial efforts to liberate the data in the first place, they think of the sanitised data as a resource for *exclusive* internal use. This is a key challenge for the open data community in India.

4.2. Accessing Government Data

Organisations surveyed in this study typically access government data through websites of various government agencies, by directly requesting them from the government offices, or by submitting Right to Information requests. Although downloading of data sets from government websites is a very common practice, almost none of the organisations mentioned downloading of data from the Open Government Data Platform. A key reason for this is that since these organisations have been collecting data on particular topics for a long period, they are most comfortable downloading such data from their original creator's website (say, budget data from Ministry of Finance <http://finmin.nic.in/>, and rainfall data from India Meteorological Department <http://www.imd.gov.in/> website). Publicly available government data come most often in closed formats (that is, PDF files and HTML tables), which an important barrier to converting downloaded data into usable data, especially for time-critical exercises. For example, Association for Democratic Reforms, which collects, compiles and analyses personal and financial information declared by electoral candidates (at municipal, state, and national levels), rely on a large data entry team to convert PDF files (for each candidate) shared by the Election Commission of India <http://eci.nic.in/eci/eci.html> into an usable dataset. Certain Ministries, especially those that have a long tradition of publishing data for analysis by external researchers, such as the Planning Commission or Reserve Bank of India, make almost all of their data available online in open formats.

Transparent Chennai <http://www.transparentchennai.com/>, which often accesses government data through RTI requests, mostly receive requested data in printed format. In very few cases, interestingly, they have also received data sets over e-mail and in open formats. Collecting data sets directly from government offices is also a common practice,

especially for organisations working with state- and local-level government agencies. Such data can come in both digital formats and hard copies. While collecting the data directly from the government offices, a few organisations mentioned that it is possible to copy the data from the computers of the agency concerned (that is, in machine-readable format) directly into USB drives or CDs. This, however, is not a general practice and is greatly dependent on the prior working relationship between the government office and the data-seeking agency concerned. Several organisations – such as Karnataka Learning Partnership, India Water Portal, and Transparent Chennai – reported that building long-term working relationships and trust with government agencies is fundamental to get access to these data sets. In several cases, such relationships become productive for both government agencies and non-government organisations, as the former makes certain data sets easily accessible for the latter, which in return offers data-based analyses and insights to the former, thus reinforcing mutual trust.

Satyarupa Shekhar of Transparent Chennai explained that it is not only crucial to build trust when working closely with government agencies, but also to ensure that capacity of government officials are developed in the process and that the practical knowledge of these officials are brought out and converted into re-usable data. For example, while mapping public toilets in the city of Chennai, Transparent Chennai team invited a group of junior engineers working at the municipal authority to approximately point out the locations of the toilet mentioned on the official list. This rough map was then validated by volunteers who visited each location, verified it, and added further information about the public toilet units. Further, Satyarupa pointed out that it is very important to recognise and appropriately use/share certain data sets that may contain sensitive data from the perspective of the government. It is possible that the data is collected by the government to identify its own infrastructural lacks so as to better design a remedial intervention.

Public sharing of such data, however, may lead to mostly non-productive criticisms of the government, without understanding that the data was collected in the first place so that the issue can be addressed (by the government). Gautam John of Karnataka Learning Partnership also referred to this fear of government officials that if opened up, data will be as "sticks" to criticise their activities. Given such a context, Satyarupa suggested that the intermediary organisation must ensure that the government agencies it is working with have a clear idea of the potential usages of the shared data, and that the former will not produce analyses and other usages of that data that shows the government agencies in poor light, or attracts non-constructive criticism.

Conversely, organisations like the ASER Centre face another kind of problem as government agencies do not collect certain data sets to begin with (such as, qualitative data for primary education). This necessitates ASER Centre to undertake collection of primary data by itself. P.K. Bhattacharya of The Energy and Resources Institute also highlighted the problem of lacking regularity and completeness of government data. Overall there is a general feeling that the government has failed to revise and expand its statistical machinery, especially in the face of new technologies of collecting and managing data and increasing demand for government data from policy researchers and development practitioners. Several organisations interacted with in this study are actively involved in bridging such data gaps – either through collection of primary data, or collating data sets from various (public and private) sources, or sanitising data published by government agencies, or sharing analysed data with media houses and citizens' groups. Not all such organisations, however, embrace the open data agenda fully and adopt a wait-and-watch stance towards it. This is creating insufficiency and sectoral-imbalance in demand for open government data from the non-governmental sphere, as well as reducing mediated access (through re-sharing) to open government data.

4.3. Sharing Government Data

Majority of the organisations interacted with in this study share government data in the form of various data products – analytical briefs, detailed reports, infographics for print media, online visualisations, and printed materials shared with various audience groups. However, very few of these organisations also share the sanitised and reorganised version of the data (done by the agency concerned) in disaggregated form. For organisations working in certain sectors, such as analysis of budget and governance expenditure data, re-sharing of data is much lesser a concern since the original data (published by government) is often available in a directly usable format, and the challenging task is not *sanitisation* of data but its *analysis*. Conversely, organisations working in sectors where official data is produced by multiple government agencies and are not published in an uniform and easily-accessible manner, such as India Water Portal and Karnataka Learning Partnership, it becomes crucial to not only share the analytical findings from the government data but also the collated and sanitised data itself. Again for organisations like ASER Centre and India Biodiversity Portal – that respectively collect and share data on quality of primary education and various government schemes, and locational and taxonomic data about flora and fauna – data sharing is part of their core activities, since the government itself produce little or no (publicly available) data on these topics. Organisations whose work involve collection of substantial amount of narrative and information responses from the government through Right to Information requests – such as Accountability Initiative and Transparent Chennai – face great technical hurdles in re-sharing such information and data in accessible digital formats, as those responses are usually provided in printed format. These organisations sometimes scan and upload entire paper documents for public access but this is clearly not the most desirable solution.

Though the Election Commission of India make public the affidavit documents –
containing information on social, economic, and criminal records – of the political
candidates submitted before election, Association for Democratic Reforms plays a
fundamental role in converting the scanned PDF copies of those affidavits (originally
submitted as paper documents) into actually usable data. It is, however, troubling that a
few organisations – such as Association for Democratic Reforms and ASER Centre – who
undertake a significant self-collection of data, and or liberation of data from available-
but-closed government documents, decide to not publicly share the final data (collected or
liberated) in disaggregated form. Both these organisations raise concerns regarding the
capacities and motivations of potential users of the disaggregated data as their primary
reason for not sharing the data. They do consider and undertake sharing of such data
directly with mostly academic researchers, but only customised data reports and analyses
are shared with local and regional outreach and intervention partners. Two other reasons
often given by some of the organisations that tend to not re-share, or partially re-share,
the raw re-organised data are that (a) they do not experience demand from disaggregated
data (as opposed to data reports) from from researchers, media persons, and other
individuals and organisations, and hence do not invest organisational time sharing it,
and (b) the lack of an organisational history of re-sharing data leads to a *by-default* not
sharing of the same, or sharing of the same in aggregated or closed formats.

In this context, Karnataka Learning Partnership is clearly one of the leaders among open
government data intermediary organisations in India. It not only gathers various types of
data produced by multiple government and non-government agencies, it re-shares them
across various kinds of data products ranging from downloadable dumps of the
disaggregated data, online map-based data browsing interface, to printed report cards
created specifically for various stakeholder groups. The India Biodiversity Portal must

also be mentioned here for establishing a most remarkable archiving and re-sharing facility for ecological data created by a multitude of non-government sources. Facing a great difficulty in gathering data from official/governmental sources regarding biodiversity in India, the Portal, led by researchers from Strand Life Sciences <http://www.strandls.com/> and Ashoka Trust for Research in Ecology and the Environment <http://www.atree.org/>, began collecting spatial and taxonomic data about habitation of species – flora and fauna – across India. It works with a range of partner (civil society) organisations and crowd-sources data through citizen science initiatives such as MigrantWatch <http://www.migrantwatch.in/>. While the Portal was recognised by the National Knowledge Commission as one of the five national knowledge portals, it has functioned almost fully without financial or operational support from the government. Unfortunately, it has neither been able to induce a positive response from the government in terms of opening up government data about the topics relevant. The government has established a similar initiative called the Indian Biodiversity Information System (IBIS) <http://www.indianbiodiversity.org/> – led by the Institute of Agricultural Science, Bangalore and National Remote Sensing Authority – which has neither opened up their biodiversity datasets. This highlights the difficulties of non-government efforts to aggregate and re-share open data at the face of governmental indifference, and especially the productive role sector-specific data disclosure policies must play to complement and empower such citizen/researcher-driven efforts.

Suman Bhattacharjea of ASER Centre narrated how lack of data skills and capacities of researchers and activists at large has driven ASER Centre to design courses of varying duration suitable for different kinds of organisations. These courses consisting lessons in arithmetics, statistics, collecting and analysing data, and making decisions based upon that; and are offered to government agencies at sub-district level, district-level colleges

and regional universities, field-level non-government organisations, etc. Suman explained that such kinds of capacity building efforts are critical not only to ensure effective and sustained re-usage of the data collected and shared by ASER Centre, but also to work towards a wider culture of data/evidence-based planning and decision-making within government and non-government agencies alike. Kiran Pandey of India Environment Portal suggested that existing linkages that organisations like Centre for Science and Environment has, on one hand, with government officials who produce and manage production of data, and with field-level non-government organisations, on the other, can be strategically used to strengthen and integrate the open data agenda on both the data production and consumption side. Effective and productive sharing of government data requires both publication of reliable and regularly updated data, and capacity building of potential re-users of data. Kiran argued that collaborative efforts among government agencies and intermediary organisations are vital. For example, Centre for Science and Environment organises training sessions on managing data, among other topics, for resource regulation (government) agencies and also for non-government organisations working in similar sectors. These training sessions can be very valuable sites of collaboration with the NDSAP-PMU team so that the workshop participants coming from both government agencies and non-government organisations can be introduced to the idea of open data, and also to the specific data sharing procedures and standards implemented by the Open Government Data Platform of India. Workshops like these can also be useful to encourage and train non-government organisations to open up data sets collected by them, using data standards comparable with the Platform.

It is crucial to note here that the organisations surveyed in this study only consider legal concerns as a factor preventing data re-sharing when the data has been bought from one of the government agencies (or from a commercial re-seller of government data), but not if

the data was downloaded from a government website. Data available in the Internet is automatically considered to be "open" – not only for use, but also for re-sharing. The organisations are very much aware that these data sets (available on the Internet) often come with no or vague license details but do not consider that as a problem in actuality. Importantly, data and information collected through Right to Information requests, or shared through the Open Government Data Platform (that is under NDSAP), or taken from government websites are all popularly seen as open data and information, irrespective of their potentially different legal status and licenses. It is also to be noted that several high-value datasets created by government agencies are sold as data products – such as rainfall data, physical and political maps, and Census data. It is not at all clear at what levels of aggregation the re-sharing of such data is allowed by the agencies concerned, and under general legal guidelines for government data in India.

4.4. Right to Information and Open Data in India

This study only managed to scrape the surface of understanding how the Right to Information and open data agendas and communities are conversing, or not, in India. The general sense is that there exists interests on both the sides to engage in exchanges of of knowledge and collaborative initiatives, but such efforts are yet to take any visible shape. When asked if their organisation identifies itself more with the Right to Information groups or with the emerging open data community, the representatives interviewed in this study were found to side mostly with the former. The question was obviously not asked to quantitatively understand whether Right to Information groups or open data community have greater support among these organisations, but to bring forth why these organisations would decide to side with either communities. The people who mentioned that their organisation has more operational similarity with Right to Information groups

gave the following reasons:

- Because open data community is yet to become substantial, and

- Because of the scale of the problems that the organisation engage with.

While the first reason can clearly shift the other way as the open data community expands in India, the second reason offers a great insight in the scalar difference in the data/information problems addressed by Right to Information and open data agendas, and can surely inspire future research on the topic.

People who though that their organisations has more in common with the open data community provided the two following reasons:

- Because the government data concerned is already available, and

- Because the organisation primarily works with available non-government data.

These answers also are highly insightful as they reveal how practice-oriented the open data community is in India (and perhaps elsewhere too). Organisations who consider themselves close to the community are simply those that are working with a large-scale of *open data*, either created by government or non-government agencies.

Sona Mitra of Centre for Budget and Governance Accountability <http://www.cbgaindia.org/> and Chakshu Roy of PRS Legislative Research <http://www.prsindia.org/> offered two opposite views of the relationship between Right to Information and open data agendas in India, which represent very well the range of views heard about the matter in the conversations during this study. According to Sona, Right to Information and open data agenda have a certain ideological convergence, in terms of arguing for institutionalisation of public disclosure of government data and

information, but there is a divergence when it comes to practical concerns. Right to
Information agenda, Sona noted, is interested in opening up more than simply data
collected and produced as part of governmental processes. It is interested in enabling the
citizens to have a clearer understanding of the functioning of the government itself, while
data is one way of arriving at that understanding. Chakshu argued that the Right to
Information and open data agendas in India are conceptually and practically different, in
terms of how how the agendas are articulated and operationalised. He explained that
while the Right to Information agenda is implemented in a "retail model," where citizens
have to go to the government and demand a piece (or a set) of information every time
there is a need for the same, the emerging open data agenda in India is demanding a
systemic change in management of government data.

Either converging or diverging, the dynamics between Right to Information and open
data agendas in India is also at the heart of the NDSAP. The policy document sets up a
governance framework for proactive disclosure of data by Government of India by
extending upon the general mandate of Right to Information as presented in the Act of
2005. On April 15, 2013, the Department of Personnel and Training circulated office
memo specifying the guidelines for implementation of *suo motu* (that is, proactive)
disclosure of information under the Right to Information Act.[14] The Section 4 of the RTI
Act declared that proactive disclosure of government information must happen along
with reactive disclosure, but it did not specify how this aspect of the Act is to be
operationalised. This memo made it mandatory for all government agencies to declare on
the website a detailed inventory of all datasets managed by the agency concerned, with
clarification regarding their publication status (open or closed), and to publicly and
automatically share all the Right to Information requests received by the agency

---

14 Department of Personnel and Training, Government of India. 2013. *Implementation of Suo Motu Disclosure under
      Section 4 of RTI Act, 2005 – Issue of Guidelines*. Office Memo No. 1/6/2011-IR. April 15.

concerned and the respective responses sent out. This memo creates a great opportunity for close integration of NDSAP and Right to Information Act, especially since the memo is silent in describing the technical framework of content/data management system required for hosting and distributing the pro-actively published information, and this is an area where the NDSAP-PMU team has already had significant experience.

4.5. Challenges in Opening Up Government Data in India

Unsurprisingly, most organisations mentioned that the foremost problem to be solved to make government data accessible is making the data sets available online. While growth of the number of data sets hosted by Open Government Data Platform has been very impressive, these data sets offer very little granularity, as most of them contain state-level and/or annual aggregates (with a few important exceptions, such as the daily price of agriculture produces across main wholesale markets across India[15]). A great wealth of data gathered by government agencies in India, however, is not only unavailable on the Platform but, more crucially, is not made publicly accessible (either commercially or otherwise) in digital formats at all. This challenge emanates from the very data/ information reporting structure between local, state and central government agencies.

Many of such unpublished data sets are not kept out of public circulation due to any specific characteristics of the data itself, such as the data having personally identifiable information, or the data containing potentially sensitive information (from perspectives of either national security or social harmony). The prevailing reason for the non-publication of such data is often a simple lack of precedence of that data being shared by the

---

15 See: http://data.gov.in/keywords/agricultural-marketing

government agency, or lack of confidence of the agency regarding the motivation of a non-government organisation or individual's interest in accessing that data. In a detailed survey among various central government agencies, Neeta Verma and M.P. Gupta have listed a set of six key challenges that these organisations are facing when attempting opening up and sharing of data.[16] These challenges included (a) the lack of clarity regarding the benefits (for government agencies) of opening up data, which leads to low motivation to do so, (b) insufficient capacity (human and financial) and budgetary support to streamline processes of collection, management and sharing of data, and (c) pending institutionalisation of open data within the everyday and regular activities of the government. The discussions with representatives of various organsiations in this study made it clear that what lies at the heart of these challenges within government is also the same thing that undermines an open data culture in the non-government sector – an approach to data as an end, and not as a mean to various ends.

Anecdotes from Amitangshu Acharya and Isha Parihar of Akvo, Debjani Ghosh of National Institute of Urban Affairs, and P.K. Bhattacharya of The Energy and Resources Institute illustrate how government agencies, especially the local authorities, look at data as part of the reporting procedure to agencies higher up in the bureaucratic order. Data is often produced not so much to reveal the ground realities but to resolve and negotiate inter-agency and intra-agency conflicts and interests. Further, such data exists in a deeply internally referenced manner, which makes its interpretation very difficult for anyone from outside the agency in particular, or the service delivery vertical in general. Lack of standardised database related procedures and of proper documentation, as highlighted by Gautam John of Karnataka Learning Partnership, makes this process of makings sense of disaggregated government data very time and expertise sensitive. The

---

16 Verma, Neeta and M. P. Gupta. 2013. Open Government Data: Beyond Policy & Portal, a Study in Indian context. In *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance (ICEGOV '13)* Tomasz Janowski, Jeanne Holm, and Elsa Estevez (Eds.). New York, USA: ACM. DOI=10.1145/2591888.2591949.

study suggest two potential strategies to initiate transformation towards re-engineering of internal data practices by government: (a) beginning internal usage of collected data within government agencies from the bottom-most bureaucratic levels, and (b) creating forums for direct interactions between government agencies that collect and manage data, and non-governmental organisations and individuals that would use such data.

The first strategy leads to treatment of government data by its original creators as something that is directly beneficial to the creator's own activities, and not only as things to share with higher-order agencies. Many problems with government data shared during the interviews reveal a systemic lack of capacity of higher-level agencies to ensure quality of data generated by lower-level agencies, and lack of motivation for lower-level agencies to produce reliable data. Self-usage of collected data, that is usage of data to drive internal activities, is expected to incentivise self-correction of sub-optimal and mis-guided data collection practices at all levels of the government, especially the field offices. Accountability of such data-driven governmental activities, of course, must be ensured through opening up the data for public scrutiny. Further, the Management Information Systems used by government agencies to archive government data are designed for storage and generation of aggregated reports, and not for sharing the data in anonymised form. As the government agencies often do not work with anonymised disaggregated data themselves, it becomes challenging for them to share it. As NDSAP-PMU experienced, much of the information systems used by the agencies are not built for publication of data, which necessitates a human layer for deciding, scrutinising, anonymising and submitting every set of data coming to the Open Government Data Platform.

The second strategy involves generating structured and regular interactive forums for the government agencies that produce and share data to understand and respond to data

needs of non-government organisations (including commercial ones) and opportunities and risks associated with that. The lack of interactions between government agencies and non-government users lead to misunderstandings between the two, in terms of their respective motivations and activities. As the experiences of Association for Democratic Reforms, Karnataka Learning Partnership and Transparent Chennai exemplify, a long-term relationship with government agencies can produce very effective models for data sharing. As the government agencies begin to benefit from the data analysis and information distribution undertaken by non-government re-users of data, the former are motivated to share data more easily, more regularly, and in greater quantity. Involving non-government organisations in analysing data coming out of ongoing government activities and using that to identify critical issues that require immediate attention may also help overcome the capacity shortfall (in short-run) within the government sphere to undertake data-driven policy implementation, and will also support wider creation of culture and skills in using data to engage with practical problems of governance. The potential threat created by such agency-to-agency channels of opening up government data, however, is that the actors on both sides of that relationship automatically gain a "gatekeeper" status, even if the organisation consciously avoids acting upon such priviledges. It must be noted that the lack of direct interactions with potential re-users of data impedes data sharing practices of non-governmental organisations, as mentioned above, as much as those of government agencies. The critical need, hence, is then to organise forums for interactions between data producers and data users – where both the sides can include both government and non-government agencies.

A key finding of the study was that most of the organisations surveyed here understand the term "open data" as publicly available government data, most often through the websites of government agencies, without any specific technical (that is, PDF is also

understood as open) or legal (that is, data without clear licensing details) understanding. This not only implies that the open data agenda needs to go deeper and wider among the research and advocacy organisations working with national-scale government data in India, but also reflects the present state of understanding and utilisation of "data" within the same circuit. The long experience of working with closed government data has perhaps brought down the general level of expectation when it comes to accessing, using and sharing government data. K. Srinivasalu of Association for Democratic Reforms stated it clearly that he does not mind when the Election Commission of India published affidavits of political candidates as scanned copy of paper documents turned into PDF files, and in fact finds it not "fair" to ask Election Commission of India to ensure publication of the same in machine-readable format. Given the key need for cost-benefit analysis, as highlighted here, when talking about upgrading the data reporting systems in a resource constrained situation like in India, it doubly becomes important that the open data agenda is not formulated as a stand alone project of the government, but as an integral part of the overall administrative and electronic governance reforms towards more *open* government practices.

Finally, another crucial concern was voiced in several interviews during this study. Many of the organisations working with national-scale government data are largely populated by researchers coming from social science disciplines and they are mostly trained in statistical and econometric methods for working with such data. The programming-biased conversations both within open data community meetings and in events and outreach initiatives launched by the NDSAP-PMU tend to marginalise those researchers not trained to approach data questions in terms of JSON files[17] and APIs.[18] The same

---

17 JavaScript Object Notation (JSON) is a data file format, commonly used to store data for easy reading and writing by both humans and machines.
18 Application Programming Interface (API) enables a software to be interacted with by another software through a set of predefined commands, routines and protocols.

people who cited this reason – too much programming terms – for their organisations being in the *sidelines* of the open data community meetings have always stated better skills for data analysis and visualisation as the highest priority requirement for the organisation to raise their ability to access, use, share, and communicate data and information. These organisations, thus, recognise it very well that programming skills, and those of visualisation softwares, are becoming barriers for them to participate in both conversations around (open) government data and in effectively utilising the same data. Hence, creating avenues for development of skill of existing researchers and policy advocates, and employing or collaborating with those from complementary knowledge backgrounds – such as programmers specialising in working with data, and GIS experts – are gradually becoming commonplace in India. This is a greatly heartening trend but its institutionalisation is also crucial. While data skills must find its way into syllabus of social science, public policy and research courses, concerns about public implications of data must also be integrated into programming, statistics and technical courses.

5. Conclusion

To repeat the above mentioned methodological clarification, this was *not* an effort to locate the "data intermediary organisations" in India but to survey a selected a cross-sector set of non-governmental and non-commercial organisations that enable other organisations to better work with (access, use, and share) government and non-government data, and see what opportunities and challenges they experience in (inter)mediating access and use of data and information. The study involved speaking to the representatives of these organisations and understanding their actual practices of accessing, using, and re-sharing data – primarily data published by government, but also self-collected data and those produce by non-government organisations. Three other key

issues that emerged during the course of the discussions are: (1) the nature of the "open
data community" in India and the participation of the organisation concerned in the
same, (2) the many challenges of effectively opening up government data in India, and (3)
the implications, applications, and communities around the Right to Information Act, and
the National Data Sharing and Accessibility Policy. While this study could analyse the
first two of these issues, the last one remains to be explored in details.

The study foregrounds three critical characteristics of the practices of accessing, using,
and sharing government and non-government data in India. Firstly, the term "open
government data" (or "open data" in general) is not commonly used among most of the
organisations in India that work with national-scale government data. Meaning of the
term is generally understood as data that is publicly available, either from government
websites or from government publications, without a very strict notion of the legal status
(such as, licensed for re-use and re-sharing or not) or technical qualities (such as,
machine-readable or not) of the data concerned. This, however, is not to be interpreted as
indicative of the limited engagements with government data by these organisations. In a
generally data-scarce situation – that is, a situation where scarcity of publicly available
government data is more the norm than the exception – these organisations (some of
which have been surveyed in this study) not only significantly use government data, but
undertake a range of data practices to liberate government data from non-machine-
readable reports and information sheets, to sanitise and compiles government data sets
available from different sources and not-directly-comparable forms, to make sense of
under-documented government data through carrying out interviews with field-level
government officials and others, etc. It is in this context that one should understand the
expansive use of the term "open government data" in India, where any publicly available
government data is seen, consumed and re-used as "open data" for all practical purposes.

Secondly, same kind of challenges exist within government and non-government agencies when it comes to re-sharing government data. The key problem faced by both these categories of agencies is that they have come to treat data as a product and not as a process. For government agencies, data is part of the internal reporting and communication processes vis-a-vis other government agencies or the higher offices of the same government agency. Government data is thus often produced as a response to, and as rendered necessary by, the demands of such inter-agency and intra-agency communications and bureaucratic negotiations. Similarly for non-government organisations, data is often part of the reporting and communication processes vis-a-vis the donors that fund the activities of the organisation concerned. The entire life-cycle of data within these non-government organisations are hence determined by the demands of project monitoring and evidence gathering as defined by the donor organisations. Further, this approach to data as a product implies that both these categories of agencies see collected and produced data as something that once shared would either become less valuable for the agency concerned (since other agencies will extract the value out of it), or would be used to study and criticise the activities of the agency concerned. Also, internal and/or closed circulation of the data sets lead to insufficient documentation of data collection and analysis procedures as those are intended to be used only within a predefined set of users. This makes the data difficult to interpret even if it is opened up. A critical challenge for the open data agenda and community in India, hence, is to establish open data practices as an integral process of functioning of these government and non-government agencies. Such re-engineering of data practices towards treating collected data as a continuous and shared resource for better internal and external monitoring of activities, recording of evidences, and gathering of feedbacks will not only require demand for more open data from potential re-users, but also government-wide reforms and incentive structures, and strategic interventions from donor organisations.

Thirdly, another key barrier for the open data agenda in India is the lack of organised and publicly articulated demand for open (disaggregated, updated, and anonymised) government and non-government data. While several organisations surveyed in the study enjoy a close and trusted working relationship with government agencies that allows them to easily access data held by these agencies, uncertainty regarding motivations and capacities of working with data was often mentioned as a key reason for not sharing data by both government and non-government organisations. The person-to-person relationships between data producing agencies and potential re-users, thus, critically mediate flow of both government and non-government data. Such relationships can facilitate opening up of data bypassing the formal procedures, but may also become bottlenecks. The challenge, hence, is to organise data users interested in accessing data created and shared by both government and non-government organisations. Unless such forums are being organised, perhaps through the efforts of an "open data community," the various users of open government data will continue to face the current challenges: (a) lack of a common platform for discussing data availability, (b) severe standards, documentations and quality problems with government data, (c) almost complete unavailability of data produced by non-government organisations, including academic, advocacy and commercial agencies, (d) lack of augmentation of government-published data through sharing of value-added disaggregated data by non-governmental organisations; and (e) lack of an understanding of mutual data needs and competences that continue to reinforce the existing divides between government and non-government organisations. Consolidating the varied data demands of such organisations, identifying sector-specific standards and practices of collecting and sharing data, connecting the organisations that can potentially collaborate and/or learn from each other's abilities, and actively overseeing the implementation of the NDSAP are, hence, some of the key functions that the open data community in India can take responsibility for.

6. Resources

6.1. Timeline of Open Government Data Activities in India

Access the timeline here: http://ajantriks.github.io/oddc/resources/timeline.html

Access the timeline data here: http://bit.ly/ogd_india_timeline

6.2. Mapping of Data Practices

Access the data practices table of the surveyed organisations here:

http://ajantriks.github.io/oddc/resources/mapping_data_practices.html

Access the entire data practices table in open format: http://bit.ly/ogd_india_mapping

6.3. Survey Questionnaire Document

Access the questionnaire here: http://bit.ly/ogd_india_questionnaire