The Energy and Resources Institute

PROJECT BRIEF JULY 2014

Open Government Data (OGD) and energy resources in India

The last decade in India has been an important and eventful decade for access to information and data. The Right to Information Act, 2005 is a landmark movement in the history of access to information in India. The last few years have also witnessed some initiatives to open and share data such as the National Data Sharing and Accessibility Policy of 2012 and the Open Data Portal of Government of India.

TERI, with support from IDRC and the World Wide Web Foundation, has undertaken a study to examine the availability and accessibility of Open Government Data (OGD) for improved governance of extractive energy industries in India. The study focussed on upstream coal and oil & gas sectors and examined issues faced by users in accessing data and government agencies in providing data. The study also explored how governance deficits can be managed with improved data availability.

Open government data

Government Data is defined as 'data and information produced or commissioned by government or government controlled entities'. (OKF, 2012) "Open refers to 'open access' that implies data must be accessible freely online, available without technical restrictions to re-use and provided under open access license that allows the data to be reused without limitation". (OKF, 2012) The parameters on which openness of a dataset is assessed include completeness, primacy, timeliness, ease of physical and electronic access , machine readability, non-discrimination, use of commonly owned standards, licensing, permanence and usage costs . (Sunlight Foundation, 2010) In India, several laws and policies mandate collection of data at various levels. The National Data Sharing and Accessibility Policy, 2012 (NDSAP) provides for sharing of government data. The main objective of this policy is to allow access to government owned sharable data in machine readable form through a network across the country, in a proactive and periodically updatable manner, within the framework of various related policies, acts and rules of Government of India. In the Indian context, NDSAP broadens the global OGD definition and includes human readability.

Funding for this work has been provided through the World Wide Web Foundation 'Exploring the Emerging Impacts of Open Data in Developing Countries' research project, supported by grant 107075 from Canada's International Development Research Centre. Find out more at www.opendataresearch.org/emergingimpacts

Although NDSAP was formulated in 2012, the data landscape and regime in the country have evolved over time. Some of the main developments in the history of information and data in India are given in box 1.

While the NDSAP is a good start, it is too early to assess its success in 'opening up the government'. Being a policy, NDSAP cannot mandate data sharing like a statute. It depends on various ministries and departments to provide data as they may deem suitable. A lot of data is managed at the state level and there needs to be greater engagement with state level agencies for data collection. The open data portal of India – http://data.gov.in is evolving in terms of content and design and so far only limited data on energy is available.[1]

Box 1: Major milestones in information and data in India

1923	Official Secrets Act, 1923
1953	Collection of Statistics Act, 1953
1960	Mineral Concession Rules, 1960 (pursuant to MMDRA, 1957)
1973	Coal Mines (Nationalisation) Act, 1973
1975	Coal Mines (Conservation & Development) Act, 1974
1975	State of UP v. Raj Narain (People have a right to know about public act by public functionaries)
1976	National Informatics Centre
1982	SP Gupta v. Union of India (citizen's right to Know)
1993	Directorate General of Hydrocarbons
1994	First Public hearing by Mazdoor Kisan Shakti Sangathan
1999	New Exploration Licensing Policy (NELP)
2002	Petroleum Planning & Analysis Cell (PPAC)
2002	Freedom of Information Act, 2002
2005	Right to Information Act, 2005
2008	Collection of Statistics Act, 2008
2012	National Data Sharing and Accessibility Policy (NDSAP)
2009	Dissemination and Data Policy of GSI (Ministry of Mine's Letter dates 5 June 2009)
2013	Open Government Data Platform (data.gov.in)
2014	Data sharing and Accessibility Policy of Geological Survey of India
2014	Policy for Geo-Scientific Data Generation for Hydrocarbons

*Information and data related

*Energy related

[1] As of April 2014

Key players in data

There are several institutions involved in making information and data available and accessible to the public. In our study, these were categorised into three broad categories - data providers, information providers and data intermediaries. Data providers are those agencies, which are the designated or nodal agencies responsible for collecting and publishing data. They may be sourcing data from companies or agencies, but are responsible for releasing the data officially for their domain. Information providers are the organizations that enable access to information and not just data. Their task is not to collect or publish data but provide information as and when sought. Data intermediaries are one of the most important categories of institutions as they compile existing government data with different government agencies and publish it in the form of reports, compendia, or raw data. Figure 1 shows how these three categories of stakeholders comprise the data setting in the coal and oil and gas sector in India. In terms of open access and open government in India, intermediaries play an important role as they help in collating, streamlining and publishing data. With the emphasis on machine readability and formats, this role has been strengthened and agencies with the requisite technical know-how can facilitate data release in an open, usable and manipulatable format.

Data landscape in energy sector in India

For analysing the state of data in India's coal and oil & gas sector, data was classified as economic, technical & process, environmental and social. Economic data included data on sales, revenue, trade, prices, operational costs, contributions to exchequer, subsidies, production, consumption, imports, exports, etc. Technical and process oriented data included data on resources and reserves, feasibility studies, etc. Environmental data included data on forest diversion, forest clearances, impacts on environment in terms of air and water pollution, land degradation, conjunctive use of resources, etc. Social data included data on impacts on people including issues of safety, relocation and rehabilitation, displacement, etc.

The main data providers in coal and oil & gas sectors are Geological Survey of India (GSI), Coal Controller's Organization (CCO), Directorate General

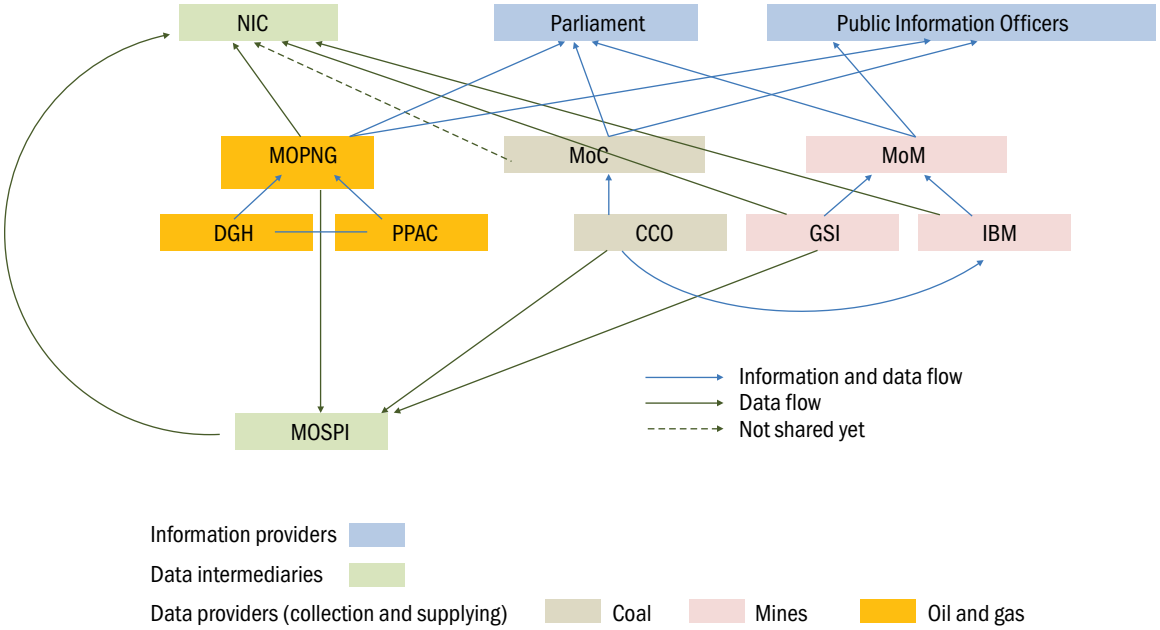

Figure 1: Institutional arrangement for information and data flow in coal and oil & gas sector
Source: TERI compilation

Data Providers	Data Intermediaries	Information providers	Users
▪ Geological Survey of India (GSI) ▪ Coal Controller (CCO) ▪ Directorate General of Hydrocarbons (DGH) ▪ Petroleum Planning and Analysis Cell (PPAC) ▪ Economics and Statistics division, Ministry of Petroleum and Natural Gas (MoPNG) ▪ Ministry of Coal (MoC) ▪ Ministry of Mines (MoM) ▪ Coal India Ltd (CIL) and Central Mine Planning & Design Institute Limited (CMPDI) ▪ Indian Bureau of Mines (IBM) ▪ State Level Departments ▪ Other Departments (MoEF, PCBs ,etc)	▪ National Informatics Centre (NIC) ▪ Data Controllers/ Managers in the nodal ministries ▪ Ministry of Statistics and Programme Implementation (MOSPI) ▪ Non-government and business organisations	▪ Public Information Officers (PIO) ▪ Lok Sabha and Rajya Sabha Secretariat	▪ Government ▪ Business and industry ▪ Civil society and citizens ▪ Academia

of Hydrocarbons (DGH), Petroleum Planning and Analysis Cell (PPAC), Economics and Statistics division, Ministry of Petroleum and Natural Gas (MoPNG), and State level departments. The main data intermediaries include National Informatics Centre, data controllers/ managers in the nodal ministries and the Ministry of Statistics and Programme Implementation. Most agencies for collection and managing data for coal and oil & gas are under statutory authority to act as the nodal agency for the data they are dealing with. The mandate and capacity of these data providers varies substantially across departments. And not all data providers have a clear mandate to distribute or share data.

The availability, accessibility, quality and openness of data vary across parameters. In the energy resource sector physical and monetary data are easier to obtain and access as compared to environmental and social data. In terms of data on production and revenue, both coal and oil sectors provide sufficient data. Disaggregated data (mine wise in case of coal and field wise in petroleum), though collected is not made available in the public domain. Overall, data in the petroleum and natural gas sector is more recent, available from multiple sources and often in 'machine readable formats' (MS Excel). Coal data is predominately provided by a single agency – the Coal Controller – is relatively more dated and available

in hardcopy and CD/ PDF in the form of Coal Directory and/or Provisional Coal Statistics.

Data portal of the government of India, http://data.gov.in, provides some energy related data, including data on upstream coal, and oil and gas. Ministry of Petroleum and Natural gas has uploaded petroleum related data on this portal while the Ministry of Coal is yet to upload coal data. However, some coal related data has been uploaded by other government agencies such as Ministry of Statistics and Programme Implementation and the Planning Commission.

Issues in navigating data

Different levels and departments of government collect, compile, and publish data. While some progress has been made in recent years to open up government data in energy resources, accessing and using this data is still a challenge.

Fragmentation: While a lot of data can be found with various government agencies, it is scattered across departments. Cataloguing of data, along with interlinking, is inadequate. Often government officials are not aware of what data is being maintained or published by other government agencies or their own departments.

Timeliness: Users contend that by the time data is published on websites of most ministries it is already out-dated. In some sectors even annual data is not published timely. Historical and comparable data is difficult to obtain. While physical and financial data are published more regularly (monthly to annual, depending on the dataset), data on other aspects, such as environment, are released in an irregular and ad hoc manner.

Errors and discrepancy: Data users have reported instances of erroneous data provided by government departments. These errors include methodological, typographical and formatting errors. Discrepancy across data from various sources is also an issue. The discrepancy is largely attributable to different definitions on resources, parameters, units and conversion, etc. Since there is discrepancy between various data and a lack of time series data, comparing data is often a challenge.

Format and usability: Many government data is still not available in electronic or online form. Even where online data is released by government departments, it is usually in the form of portable document formats (PDF). This makes the data partially machine readable as it cannot be easily reused or manipulated. The quality of pdfs varies in such degree that not all the documents uploaded are legible.

Trust deficit: Departments feel that if provided in excel or other formats, data could be manipulated. This however stems from two factors – trust deficit and inadequate use of security systems. A lot of data is not made accessible citing security and confidentiality as a reason. However the process and classification of data into secret, confidential, classified, etc., is not clear to users.

Elite capture: Data that is accessible is not always easily comprehensible. In this context, there is a perceived risk of elite capture amongst the users. Select players with access to data and the know-how to process data can capture the domain of open data.

Constraints in opening up government data

Importance of data transparency and openness is widely acknowledged but not reflected in practice. Several challenges and impediments exist in sharing data by government agencies.

Confidentiality concerns: There are concerns on sharing data that could be confidential and their are concerns that data can be misused/ misreported. Representatives of government agencies often fear that citizens may use raw data irresponsibly and an incorrect analysis of data may lead to misconceptions and may hurt business interests or larger public interests. Hence, many data suppliers are cautious in releasing any data in the public domain. There are concerns about sensitive information as there have been past instances of confidential information finding its way into the websites and reports sold by data intermediaries.

Priority and mandate: The lack of priority given to data, its collection and dissemination is a key reason for inadequate opening up of government data. Most reactive or proactive data provision initiative in India is seen as an increase in effort without provision of appropriate resources. Often data suppliers and providers feel that there is little clarity on the need and purpose for open government data and this is not in congruence with the

requirements of intended stakeholders. Even where certain institutions have a clear mandate to collect and report data, this mandate is not clearly communicated to other agencies who do not manage data as their mandate.

Capacity: Data suppliers and providers feel that there is a lack of information infrastructure to provide machine readable, reusable and easily interpretable data to the public. While more and more policy makers are relying on data for decision making, the resources allocated to data collection and classification have not increased accordingly. Most data supplying and providing agencies lack dedicated trained statisticians and IT professionals to maintain and supply data.

Awareness: Overall, the level of awareness about the NDSAP and Open Data portal varies amongst stakeholders. While most of the Ministries concerned are aware of the initiative, many of their respective data related agencies are not. In many cases, it is only the Ministries which interact with NIC on data and the data agencies only provide data to their respective Ministries.

Coordination: There are issues of co-ordination between different government agencies, which hampers timely collection and dissemination of data. Numerous agencies collect and provide similar data. Not only do these agencies use varying definitions and assumptions, at times different formats are also used to report the same data.

Enhancing openness of government data

Based on the study and stakeholder consultations, several suggestions have emerged as possible solutions to address the gaps and challenges that exist with respect to the data landscape of the country. While our lens of study was energy resources, especially coal and oil and gas, these suggestions hold true for OGD in India in general and can be of use across sectors.

Policy measures

Fostering a culture of openness

Laws, policies, programmes, have been introduced from time to time to improve transparency in the working of the government. The success of these initiatives however varies and there is a need for stronger enforcement of existing laws and policies. More importantly there is a need to create a 'culture of accountability and transparency' in the government and there is need for further 'sensitising government officials and staff' of their 'accountability to the public'. There is also a need for greater awareness amongst citizens on their 'right to hold government accountable'. Without enhancing awareness levels, any move for transparency, including the Open Data, will remain a prerogative of the elite and urban professionals.

Making open data useful for improved governance and service delivery

An issue with open data as pointed out by (Janssen, 2012) is the need for 'intellectual accessibility'. Without the ability to interpret the vast amount of datasets made available, many citizens will not be able to make sense of the data. In such a scenario data will create more inequality as it will redistribute the source of power (through data) to another group of educated elite. There is need for 'intermediaries' to effectively utilise data, develop visualisations and applications that make sense of data. International experience shows instances where public data has been utilised by developers to come up with applications that have improved transport planning and public delivery. There is need to find ways to incentivize IT developers in India to come up with such applications. Incentives would include stricter protection of developers copy right of their applications or sufficient remuneration/revenue stream for the developer.

Improving data quality

There is need for improving quality of government data through better record keeping, improving co-ordination between government departments and reducing multiple departments and personnel providing similar data. Some specific suggestions on improving data quality are:

Independent verification

Independent verification of data and/or appropriate quality checks needs to be in place before putting out data. Where data needs to be collected and complied from different agencies, formats and processes should be such so as to require minimum human intervention. This would reduce the margin of error. Discussion with data suppliers suggest that some government

departments are moving towards more real time and machine collected data. Such practices need to become more common.

Streamlining data

Data sharing formats need to be made uniform across government departments in order to ensure inter-operability. Data users suggest that having a single source of data for 'Energy' as against multiple sources for different fuels would streamline data and make analysis and comparison more easy. USA for instance has a single source of all energy data – the Energy Information Administration (EIA). The website provides comprehensive time series data. Data is also presented in a user friendly manner and updated regularly. Such best practices can be studied to improve existing data websites in India.

Confidentiality

From stakeholder discussions, it emerges that the 'need for confidentiality' emerges as a major impediment in getting access to government data. There is however a lot of ambiguity on what data is confidential and what can be shared. The NDSAP does suggest government departments to prepare a negative list of data that cannot be made available on the public domain.. However it is not clear whether such a negative list has been prepared by various Ministries. If not, than such a list should be prepared by government departments and importantly this list should be made public so that there is clarity and transparency on what is public and what is not. In certain cases 'confidentiality of data' is also limited to a certain time period after which the data can be made publicly available. The negative list should also classify what data can be published subsequent to the 'confidentiality period'.

Definitions and methodologies

In terms of consistency in data, a common concern has been on 'definitions' and 'methodologies' on how estimates are arrived at (for instance for oil reserves, refining capacity). Discussions with users and suppliers suggest that often there is no consensus or single definition/methodology as a result of which same data from different agencies may differ. Such inconsistencies can be avoided by departments dealing with similar data by arriving at a consensus on definitions.

Improving effectiveness of data portal

Open Data Portal of the Government of India is a very good initiative, which is still evolving and experimenting but growing at a good pace. The portal has been instrumental in putting a lot of government data in machine readable format but there is a significant scope for improvement.

Awareness

Open Data Portal has been in place for almost a year and a half but awareness about the data portal is low amongst data users as well as data suppliers. Data controllers are nominated in each government Ministry but the staff, even those dealing with data, are always not aware of the kind of data that is available on this data portal. Thus, awareness needs to be built especially amongst government agencies to facilitate better contribution of data towards this portal and ensure that government itself uses this data in its decision-making. With respect to awareness amongst non-government users, many users are still relying on the data sources used earlier.

Linking to avoid multiplicity

More data does not necessarily mean more openness in government and its data. Number of datasets should not be a criterion for reviewing performance, as often similar datasets are provided by different government agencies. It is important that there is a system in place that links to different datasets that provide the same information, either on the data portal or elsewhere. Reducing multiplicity is a key step in avoiding confusion and lack of clarity about which is the most reliable and best quality data source.

Usefulness as important as usability

Open Data Portal in its short span has put a variety of data in a readily accessible and machine readable format. While NDSAP implementing agencies and the Project Unit are largely responsible for facilitating the uploading of data in open data formats, there needs to be a system in place to review the contents of data portal from time to time. While the focus on a machine readable format is immensely helpful for usability of data, it may not always be useful. Usefulness of data is subjective and would vary

depending on the sector and use. Therefore, it requires a close interaction, even a joint responsibility, with the different Ministries and Departments.

Criteria for seeking data

NDSAP and data portal depend on various government agencies to upload data. The portal allows citizens and users to place a request for a certain dataset online. Once the number of requests for a dataset crosses 100, NIC officially writes to the department concerned requesting it to provide that data. While this is a very good practice where citizens can directly influence the data availability, it is important for the data portal and its implementing agencies to seek data in a structured manner. Towards this a multi-stakeholder group can be engaged with to help in proposing a clear list of data that can be uploaded for different sectors.

References

Janssen, K. (2012). *Open Government Data and Right to Information: Opportunities and Obstacles*. Retrieved May 14, 2014, from The Journal of Community Informatics: http://ci-journal.net/index.php/ciej/article/view/952/954

OKF. (2012). *Open Government Data Handbook*. Retrieved April 30, 2014, from Open Knowledge Foundation: http://opendatahandbook.org/pdf/OpenDataHandbook.pdf

Sunlight Foundation. (2010). *Ten Principles for Opening Up Government Information*. Retrieved from Principles for Opening Up Government Information: http://assets.sunlightfoundation.com.s3.amazonaws.com/policy/papers/Ten%20Principles%20for%20Opening%20Up%20Government%20Data.pdf

CONSULTATIVE WORKSHOP ON 'OPEN GOVERNMENT DATA AND RESOURCES'

(17 JULY 2013)

The Energy and Resources Institute (TERI), with support from the World Wide Web Foundation and the International Development Research Centre (IDRC) organized a consultative workshop on 'Open Government Data and Resources' on 17th July 2013 at TERI, India Habitat Centre, New Delhi.

The workshop brought together stakeholders from the government, industry, academia, think tanks and civil society for a discussion on issues around open data in natural resources, particularly in case of coal and petroleum.

The discussion brought out perspectives from users of data, suppliers of data and the intermediaries towards understanding the state of 'government data' in these sectors and how greater openness can improve efficiency and overall governance in coal and petroleum sectors. Main themes of the workshop were Energy resources: Plugging the data gap; From Government to Citizens: Suppliers' perspective; From availability to access: Users' Perspectives.

Participants at the workshop discussed how better availability and access to data in the energy sector can help in improving the governance of coal and petroleum development. Lack of adequate reliable data can have various implications in the form of influencing policy formulation, institutional accountability, reliability of research results, and international perceptions. Several issues were raised vis-à-vis the existence of data and its adequacy and usefulness.

THE ODDC NETWORK

This study is part of a research network 'Exploring the Emerging Impacts of Open Data in Developing Countries' (ODDC). More updates on the project are available at http://www.opendataresearch.org/project/2013/teri

For more information contact:

Ms Nidhi Srivastava
Centre for Resource and Environmental Governance
The Energy and Resources Institute (TERI)
Darbari Seth Block,
IHC Complex, Lodhi Road,
New Delhi- 110003
Tel: 24682100 or 41504900
E-mail: nidhis@teri.res.in

Ms Veena Aggarwal
Centre for Resource and Environmental Governance
The Energy and Resources Institute (TERI)
Darbari Seth Block,
IHC Complex, Lodhi Road,
New Delhi- 110003
Tel: 24682100 or 41504900
Email: veenaa@teri.res.in

The Energy and Resources Institute