<div align="center">

Consultative Workshop on

Open Government Data (OGD) and resources

Venue: TERI, India Habitat Centre, Lodhi Road, New Delhi 110003, India
Date: 17 July 2013: Time: 9:30 – 13:30

Summary of proceedings

</div>

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 government data in energy resources, particularly in case of coal, petroleum and natural gas.

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.

Governance challenges in energy resources: Plugging the data gap

This session was directed towards exploring the extent to which access to data can address governance challenges in the energy resources sector. The discussions also delved into the domains of data accessibility gaps and the crucial nature of government in data provision.

There is a growing concern that the Indian economy is gradually moving towards a capture economy where the private sector has been capturing the regulatory process especially in resource based industries which are 'rent thick'. Since data frames the narrative of governance challenges it is important to ensure that the government data does not create a false or inaccurate narrative. A lack of data accessibility or opacity in data has implications for undue advantage to private firms, competitions, transparency in government actions and political accountability. In all the energy sectors, it can be seen that data shows whether there is a huge gap between demand and supply and if there is a need to infuse investment in the sector to remove this gap. The Indian energy problem is framed as a problem of scarcity of supply and this supports the narrative of giving support to the private sector, so as to increase the supply. The Indian energy problem is social, economic, and political in nature and data for this is inadequate.

Particularly in the sectors under study, it was observed that data on energy resources held by private sector is very opaque. In the petroleum sector, production figures from private firms are inadequate, complex quantitative figures produced by under-recovery - which conceal rather than reveal information, LNG import data does not state pricing formula clearly among other issues.

In the coal sector, resource data does not meet international standards, where techno-economic viability is considered. The Government consistently refused to use UNFC classification as a basis for calculation of reserves. Moreover, the resources figure have irregularities of including resources which are not extractable or which are already mined etc. Data on entities involved in resource development in the coal sector is inadequate, which becomes the basis of collusion between private sector and government.

In both the sectors there is lack of data on the extent of financial concessions to private firms. There is a lack of phase wise data for extractive industries. Exploration data is key to public interest in resources sector in order to make the right regulatory decisions.

Since the government is dependent on data from private firms for some sectors, the lack of public accessibility to data reduces the transparency and accountability in the sector. There are issues of commercial confidentiality of data and security of the country. A lot of data is not made accessible to public stating national security concerns but there is inadequate clarity on the meaning of national security. Classification of data is there in the Ministries: confidential, secret, classified etc. Ministries, in their own wisdom, may have classified into these categories at some point of time. In any case, these decisions or classifications are not well debated. Thus the classification can be questioned. There is a need to revise this classification on a periodic basis. Many a time even classified information is published by the companies. Thus there is an information gap about what is classified and what is de-classified.

From availability to accessibility: Users' perspectives

The session on users' perspectives on open government data started with the question whether policy decisions are actually taken on the basis of evidence or whether evidence is created to support policy decisions. From a users' perspective, data problems can be divided into different categories.

First category is data that is unavailable

- Coal / Oil / Natural gas: Field-wise reserves, production data is difficult to get. No credible estimates for the demand for natural gas in India. No comprehensive data on gas off take by refineries, captive etc.
- Even data on usage is very difficult to get. For instance which sector is using which and how much of energy resource - modal shares of oil use within transport sector.
- Environmental compliance information is missing and therefore data on environmental impact of energy projects is practically non-existent.
- Many projects are under an obligation to provide emissions, affluent related information and share it with the State Pollution Control Boards (SPCB), which then makes it

available online. However, such data are not being provided. Information on land, water use of energy projects is also not available.

Second category is data that is available but difficult to access

- Coal statistics is prepared and published by the Coal Controller's Office (CCO). But in practice, the CCO website usually has incomplete historical data and updated versions are not available. Even annual reports of CCO are available 4-5 months later. There are a lot of differences between provisional statistics by the CCO and the actual annual report. There is a feeling that meaningless data is made available to public to intentionally mislead.
- Getting data is often dependent on some individuals, not a part of the system. It is difficult to even get data that is known to be collected.
- Other examples of such data that is there but difficult to access include electricity tariffs, consumption by economic class, CEA annual electricity statistics, which is available only in print. E auction data is not available anywhere in the public domain. It is available through other priced publications.
- CEA Annual data is available only in hard copy. It is easy for people in Delhi to get access to hard copies but not for people in other cities.
- There is fragmented availability of data and use of inconsistent formats and frequencies for providing data. Central Electricity Regulatory Commission (CERC) comes out with market monitoring reports. It has data only on that month. So a comprehensive month wise data is not available.

The third category is data that is available but unfortunately erroneous

- One instance of it is in the Energy Statistics report from MOSPI. In 2012 Report there were a number of copy paste errors, where numbers were pasted across tables. There were also a few methodological errors. Numbers also do not match with other data (for example the electricity consumption data does not match with coal data)

Other problems with data availability include that of inconsistencies and data relevance. There are inconsistencies in data for instance, in the estimates of wind energy potential by C-WET and others, estimates of emissions from power sector differ by ~15%; and inconsistencies in activity data and energy use data for transport. In terms of relevance of data, there is a lag in publication of data, usually of at least 2-3 years. In social sector data, there are large gaps between multiple rounds. For example, surveys such as National Family Health Survey (NFHS), HDI reports of states. Even when the states did submit these reports, multiple states used different methodologies and their reports are not comparable with each other. Even when data is published, one does not know if it is valid, useful etc. There is of course an issue of coordination amongst different agencies to ensure consistency of data.

As the RTI is becoming a popular tool for data requisitions, the government is gradually removing data from their websites and encouraging people to submit RTIs. A lot of data is also presented for parliamentary questions which is not made available otherwise.

In the petroleum sector, the Petroleum Planning and Analysis Cell (PPAC) has two links – historic and current. On historic, they are giving data up to 2011-12 and the current data provides 2013 - 14 so 2012-13 data is missing.

The coal sector is very opaque in terms of data accessibility. In terms of resources and reserves of coal, the terms have been used interchangeably as per willingness by institutions such as the CMPDI. The coal Directory figures have been same for several years implying no extraction has happened.

Researchers in both the sectors find it difficult to do any time series analysis due to lack of continuous data.

Based on these categories, there are various implications for public policy –

- Weak policy formulation – policy making in the country could have serious social, financial, environmental implications since the basis of these policies may be incorrect. Energy policy in particular needs objective information because decisions have long term implications.
- Lack of institutional accountability without publicly available data.
- Research results are faulty and questionable if based on unreliable data.
- These inconsistencies create erroneous international perceptions as well

Some of the suggestions for improving data availability include the formation of a nodal agency – housed in the planning commission or PMO – which would not be limited by a ministry's jurisdiction. The agency could be tasked with collecting, cleaning and reconciling data from multiple sources to inform energy related policy formulation and create better coordination between multiple agencies. The effective implementation of National Knowledge Network (NKN) can also be one of the important supports for policy formulation.

Within the current system, there is a need for standardized, frequent data collection across multiple agencies along with online publication of data in easily accessible formats to ensure transparency, accountability and reduction in research costs.

From government to citizens: Suppliers' perspectives

The third session was focused on the challenges faced by suppliers of data or data providers in making government data open and accessible. There are policies such as the National Data Sharing and Accessibility Policy (NDSAP) in place which is based on the premise that all government departments, ministries and agencies should release their data in an open format using the platform. NDSAP also provides for government department to come up with a negative list which lists data that is prohibited from sharing under any of the prevalent law. India does not have a privacy law as such but there is an understanding that private citizen's

information is not to be released. State governments have also been advised to adopt this policy and release their data.

As a data provider, NIC has been given the responsibility of setting up the platform and setting some standards etc. for different kinds of data. This platform will not remove any data thereby making it possible to do a time-series analysis even if the nodal department is maintaining a yearly data. Data once provided in this portal is archived for the time to come. Any data set entered goes through two stages of moderation. First, approval by the data controller/manager at the nodal department/ministry; Second, PMU at NIC which looks at format related issues.

On the data.gov.in portal, data users can download, explore, visualize data etc. it is meant to be a platform to facilitate dialogue between data users and suppliers and NIC tries and mediates between the two. Data users can place requests for data; can point out errors, rate the dataset, comment on quality, publicize on social networking sites and within sectoral communities as well.

The portal provides details of the person who has provided the data and users can write to that data controller directly. A user can request for a dataset. After 100 endorsements, NIC approaches the nodal ministry to release that data. In our study sectors, there are 12 datasets from Ministry of Petroleum and Natural gas (MoPNG) while Ministry of Coal has not contributed any dataset to the portal.

While datasets are important, promoting the use of data is also needed. Citizens should also make use of data. Towards this efforts are being made towards making such data available on phone through community contribution applications.

The PPAC collects data on behalf of the MOPNG and has been making data available to public since 2003-04 even before the NDSAP came into existence. PPAC hosts most data on downstream of the sector with little data on the upstream. Data is collected from all the oil companies, which is already a processed data coming from their auditing system. Mostly data is authentic, as it comes from the ERP system of the companies. There is no human intervention. However, if there is any inconsistency PPAC approaches the companies. Then the data is compiled and collated; verified and authenticated; report is prepared and uploaded on the PPAC website. Most of the data is uploaded on a monthly basis. Even when it is provisional, it is uploaded every month.

From a suppliers' perspective, some of the issues were identified as follows –

Issues	
Clarity on Need & Purpose	Generally not specified & not in sync with intended stakeholders' requirements
Mandation	At not times not legally supported (Unlike PPAC, which has a clear mandate; Pvt sector does not have that kind of a mandate) (No obligation on private agencies to share data)

Multiplicity of Agencies	Same data tends to get repeated & many a times is required to be re-submitted several times at different levels (Excuse: all the data is available at customs port too; so users asked to take data from there)
Resource Allocation in Data mining & submission	Separate allocation of Resources required at Suppliers' end
Level Playing Field	Generally PSUs' are made to comply- Private entities may not have obligation to share data
Commercial sensitivity	At times, data submission leads to sharing of commercially sensitive information/ data
Data integrity & updation by collecting agency	Very often not done in time & looses its very purpose- becomes a dis-incentive; Industry experts can be involved for this
Fear that Data Analysis may not be meaningful	Incorrect analysis leads to misconceptions & may hurt business interest
Data Integrity Checking mechanisms not present	Industry Experts can only disseminate data & make meaningful comparisons. Very often novices are involved in collection & analysis
Formats for data collection not scientifically designed & not appropriately communicated	Focus on purpose is lost; Request is sent to the head. By the time it trickles down to the relevant department, the meaning, purpose is lost

It is difficult to get complete business data from private companies as it may contain some business sensitive information too. PPAC does not share their business confidential data but it takes time to build that trust with the companies. One has to make sure that data is not costing as much in terms of time and money. Timeliness of data has to be ensured once it is decided that data will be uploaded on a monthly basis. Most of the data enlargement has also happened in response to RTI applications received.

There is a need to demystify prices but some of the contracts also state that information is not to be shared in the public domain. But there is information that can be shared, for example, in the case of spot prices which is not given even at the end of the year. There are issues that are entrenched into the sector like users' resistance and monopoly. Competition law is the solution

here to clarify what should be commercially sensitive and what should be brought in public space.

Illustrations		

Data on capacity in Pipelines & Re-gasification Terminals	✓ No clarity on
	
	   ▪ which is the right agency legally mandated for its collection
	
	   ▪ the purpose of its intended usage
	
	   ▪ Periodicity of its updation
	
	✓ Data is already available online with Customs & Excise authorities
		▪ PNGRB is legally mandated agency since 2006
		
		▪ Purpose is to determine & make shippers know of Common Carrier Capacity available for booking
		
		▪ No common platform for data sharing exist on PNGRB website

Data on demand of POL Products	✓ Bare minimum information is put on public portals by collecting agencies, very often late	Basis & Assumptions not shared making Data integrity & timely analysis meaningless; Sectoral experts can help in this respect
Data on All-India LPG customer base	Consistent & common basis across industry was successful in weeding out multiple domestic LPG connections & diversion	_Became Possible_ since GoI mandated oil PSUs' to check unintended subsides & share database with each other; This is where a shared database has made a difference

Some suggestions from the suppliers' side include the creations of a mandated single point national repository of databases with industry experts which will identify purpose of data collection based on stakeholders' requirements; maintain scientific methods & consistency in approach in collection, analysis & dissemination; continuously examine data integrity; create an interface with users & suppliers & balance their perspectives; and use latest technology in data collection, analysis & dissemination processes. Such an agency would require power to impose penalty and even undertake peridoc independent audit of data availability and fulfillment.