Energy Research & Social Science 25 (2017) 105–109

Contents lists available at ScienceDirect

Energy Research & Social Science

journal homepage: www.elsevier.com/locate/erss

Perspectives

Data for development: The case for an Indian energy information administration

Varun Rai[a,*], Rahul Tongia[b], Gireesh Shrimali[c], Nikit Abhyankar[d]

[a] LBJ School of Public Affairs and Department of Mechanical Engineering, The University of Texas at Austin, USA
[b] Brookings, India
[c] Steyer-Taylor Center for Energy Policy and Finance, Stanford University and Director, Climate Policy Initiative, India
[d] Lawrence Berkeley National Laboratory, USA

ARTICLE INFO

Article history:
Received 5 January 2017
Accepted 6 January 2017
Available online 19 January 2017

Keywords:
India
Energy data
Energy governance
Energy institutions

ABSTRACT

Energy in India has a big data problem – the world's second most populous and fastest growing large economy does not have a singular central body in charge of maintaining and disseminating India's energy data, let alone analyzing it. Different pockets of data come from different bodies, some statutory, but these lack granularity and timeliness, even before we get to the challenges of data completeness, accuracy, and differences in methodology and assumptions. We propose that India should create a national Energy Information Agency – an Indian EIA or "indEIA." India urgently needs a dedicated, central agency to collect, collate, disseminate, and facilitate the analysis of all essential energy-related data. An entity like indEIA will be critical in helping India leverage the creative powers of the national and international analyst community to provide reliable, cost-effective, and clean energy to its citizens. Our proposal envisions indEIA as the primary vehicle for curating and maintaining India's energy data.

Energy in India has a big data problem – the world's second most populous and fastest growing large economy does not have a singular central body in charge of maintaining and disseminating India's energy data, let alone analyzing it. Different pockets of data come from different bodies, some statutory, but these lack granularity and timeliness, even before we get to the challenges of data completeness, accuracy, and differences in methodology and assumptions. This is especially true for up-to-date data, exacerbated by the growing energy demand and new government programs that make certain data obsolete very quickly. Lack of robust datasets has been a stumbling block for serious research and analysis on energy in India, evidenced, for instance, by limited data-driven analysis in top journals. Lack of data availability also limits effective public participation in energy policymaking. For example, non-governmental analysis is ostensibly welcomed in India, but often limited in impact as policymakers have, at least in the past, primarily relied on government bodies for inputs. Limited impactful input from independent analysts in the policymaking process is exacerbated by the fact that analysts often do not have the data to undertake appropriate analysis. The real loser in the process is the Indian public, which continues to face unreliable and relatively expensive energy services [1,2].

Availability of good data is crucial for robust policy analysis, including policy design and evaluation and for enhancing transparency in energy sector governance [3]. This is especially true in a country like India that needs to rapidly expand its energy infrastructure, improve the availability, affordability, and reliability of energy supply to its masses, and has a leapfrogging opportunity in doing so. To realize this potential and opportunity, however, will require leveraging appropriate policy tools complemented with careful planning, implementation, evaluation, and course-correction as learning accrues over time [4]. As India embarks on its critically vital mission to expand and modernize its energy infrastructure and supply, it needs to leverage energy data as a national asset.

We propose that India should create a national Energy Information Agency – an Indian EIA or "indEIA." India urgently needs a dedicated, central agency to collect, collate, disseminate, and facilitate the analysis of all essential energy-related data. There are a number of government support programs, for everything from renewable energy to electricity access, in the billions of dollars. Their efficacy is rarely understood, in part because of data limitations. An entity like indEIA will be critical in helping India leverage the creative powers of the national and international research community to provide reliable, cost-effective, and clean energy to its

* Corresponding author.
E-mail address: raivarun@utexas.edu (V. Rai).

http://dx.doi.org/10.1016/j.erss.2017.01.002
2214-6296/

citizens. In addition, granular data can enable the industry as well as policymakers to move towards more customized, dynamic, and nimble solutions, instead of relying on averaged, aggregated, and non-granular time-series or spatial data. Thus our proposal envisions indEIA as the primary vehicle for curating and maintaining India's energy data, a vital national (and global) asset.

Before getting too far into our commentary, we note that we draw relevant examples from the U.S. for informing the roles and opportunities for indEIA. To be clear, though, we are not suggesting that the U.S. is the right institutional context to base indEIA upon. Rather, our choice of examples from the U.S. is a direct reflection of the fact that the U.S. EIA has one of the most established, expansive, transparent, and widely resorted to (by analysts and modelers globally) operations in energy data and analysis. There is much to learn from the U.S. EIA, not just about its laurels but also its pitfalls (especially in modeling and projections; see for example [5]). On the other hand, just like India countries such as Brazil, China, and South Africa also reflect conflicted institutional structures and path dependent evolution of their energy sector. But that overarching similarity also hides deep differences that the energy sectors and institutions in these countries have vis-à-vis those in India. Thus, while it is tempting to draw relevant comparisons with countries such as Brazil, China, and South Africa, what exactly comports to the Indian situation is quite context dependent.

The importance of having robust energy data is a top-priority item for the broader analyst and scholarly communities and is an issue now commonly discussed in stakeholder exchanges. For example, a number of scholars (for example, see [6]) have called for an integrated energy modeling institution, a role that the erstwhile Planning Commission (now reborn as the *NITI Aayog*) was meant to cover, such as through its Integrated Energy Security Scenarios. NITI Aayog, in fact, hosted a multi-stakeholder workshop on this topic on August 10, 2016, when they also signed cooperation agreements with the US EIA and UK on energy data and modeling. Such modeling by Planning entities is widespread, including by China's National Development and Reform Commission. However, to enable wider analyses and independent modeling by multiple entities and researchers, we believe the first step should be data gathering, quality-control and validation, and dissemination. This helps not just modeling but also awareness, compliance, replication, and checks and balances.

To be sure, there are worthy ongoing efforts by both governmental and non-governmental bodies for energy data. In 2012, the Government of India launched *data.gov.in*, an online platform where multiple ministries (voluntarily) share relevant datasets in the public domain. Although it is a step in the right direction for integrating data collection, its effectiveness is limited as yet because it only collates the already available data, which at times are not up-to-date, lack granularity, and not organized in a user friendly manner. Similarly, the Central Statistics Office in the Ministry of Statistics and Programme Implementation (MOPSI) publishes an annual "Energy Statistics" volume that compiles data on multiple sub-sectors like natural gas, power etc. In the oil and gas sector, one of the central functions of the Petroleum Planning and Analysis Cell (PPAC) is data collection and analysis. In the power sector, the Central Electricity Authority (CEA) publishes extensive data in the electricity sector, but some are not up-to-date, and are often heavily assumption driven, relying on states reporting the data. Similarly, the Power Finance Corporation (PFC), a public-sector firm, publishes annual reports on the working of state distribution utilities. Thanks to the state electricity regulatory commissions, utilities in a few states have started sharing rich data on electricity usage and costs in the public domain. For example, the Maharashtra State Electricity Distribution Company Limited (MSEDCL), the state distribution utility, till recently shared hourly loading of all its 11 kV consumer distribution feeders on their website (although in PDF format). PFC is creating a National Power Portal to bring together data from states, leveraging major information and communications technology (ICT) programs on metering [7]. The National Sample Survey collects extremely valuable information on consumption through surveys; however, their energy consumption data are insufficient for policymaking. In the private sector, in 2015 The Energy Resources Institute (TERI) published its 29th annual database of energy publication, TEDDY. A good example of a collaborative effort between India and the U.S. is the Indian Renewable Energy and Energy Efficiency Policy Database [8], which was motivated by the success of Database of State Incentives for Renewables and Efficiency (DSIRE) in the U.S. Finally, thanks to the Internet, there is more energy data on India online than ever before. However, those data are not standardized, let alone the veracity of underlying methodologies and assumptions used to compile them.

In spite of these efforts, there are a number of shortcomings, especially regarding comprehensiveness and timeliness. One obvious issue is that the data are in silos, sometimes by Ministry. Missing are linkages across energy domains, let alone to allied Ministries, such as transportation. Data are often out-of-date and not granular enough. For example, official CEA data for "shortfall of power" are not granular by time of day, making it difficult to assess the optimal electricity generation mix to serve the load [9]. Central data collection efforts, such as by the PFC, often rely on either "nodal agencies" that collect data from the states or the states themselves for many of the data. In fact, the reliance of most of these efforts on the states for primary data is one of the bottlenecks in the system. Ideally, data should be not just digital but in machine-readable format. Ministry of Power's recent (2014-15) annual report is online indeed, but the publicly available PDF version is a scanned version of the report, requiring laborious data-transcription efforts by scholars trying to analyze the data. Moreover, the challenge isn't only about the quality and dissemination of data – often the data just aren't available, e.g., a significant fraction of electricity feeders isn't instrumented, though online real-time monitoring programs by the Ministry of Power such as *Vidyut Pravah* and *Urja* are trying to address such shortcomings.

Lack of good data seriously hampers robust policy analysis, which requires much more granular (both temporally and spatially) and consistent data than are available from official sources in India. For instance, besides broad descriptions and macro statistics, there is little public, collated record of the scale and scope of support the government offered to individual RE programs, hampering empirical studies analyzing the effectiveness and efficiency of the various support schemes. A good example is a recent paper by one of the coauthors examining government support for wind energy in India and the effectiveness of different incentive schemes (*reference not shown to comply with double blind requirements*). The analysis required plant-level performance data, which was not easily available in India, leading to the authors using a less than optimal (i.e., small) set of data from the clean development mechanism (CDM) database.

Data issues in India's solar sector are also staggering and especially telling since solar is at the forefront of the central government's push to address a myriad of energy needs. There is all-round paucity of data on the deployment of solar energy in India (for example, see [10]). There is no authentic, validated dataset that provides even a reasonable picture of the evolution of installed prices, technologies, market structure, and policies (local, state, and central) for solar. Scholars often fall back to commercially available databases, which inherently suffer from validation issues. In fact, even the Ministry of New and Renewable Energy (MNRE) does not directly generate (or disseminate) data for the annual generation of power from RE, relying on CEA for such data, which, in turn, relies on the states to try and capture such data. Unofficial accounts (online blogs, discussion forums, etc.) are replete with discussions