A thesis submitted to the Department of Environmental Sciences and Policy of Central European University in part fulfilment of the

Degree of Master of Science

Data is the New Fuel: Challenges to Energy Data Management in India

mridul PAREEK

July, 2019

Budapest

Notes on copyright and the ownership of intellectual property rights:

(1) Copyright in text of this thesis rests with the Author. Copies (by any process) either in full, or of extracts, may be made only in accordance with instructions given by the Author and lodged in the Central European University Library. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the Author.

(2) The ownership of any intellectual property rights which may be described in this thesis is vested in the Central European University, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement.

(3) For bibliographic and reference purposes this thesis should be referred to as:

Pareek, Mridul. 2019. *Data is the New Fuel: Challenges to Energy Data Management in India.* Master of Science thesis, Central European University, Budapest.

Further information on the conditions under which disclosures and exploitation may take place is available from the Head of the Department of Environmental Sciences and Policy, Central European University.

Author's declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

mridul PAREEK

CENTRAL EUROPEAN UNIVERSITY

ABSTRACT OF THESIS submitted by:

mridul PAREEK

for the degree of Master of Science and entitled: Data is the New Fuel: Challenges to Energy Data Management in India

Month and Year of submission: July, 2019.

Energy data is central to the development of knowledge, new business conducted, and governance enacted around the energy sector. Because of these benefits, existence of well-developed energy data management in a country is very important. Lack of data and data collection mechanisms hamper multifaceted research and planning activities which depend on it. In a rapidly growing country like India, the scarcity and inaccessibility of useful energy data brings more harm and has been criticized by scholars as a bottleneck in the system. Studies, particularly qualitative, on this topic are very limited. Studies around energy data are predominantly about the economics and performance of schemes and are built upon technical parameters. The research aims to address the knowledge gap by investigating the challenges to the existing energy data management system in India. Through reviewing literature and semi-structured interviews with experts in the field, the research points to four dominant themes. These are: the existing collection mechanism itself, present data gaps, socio-political motives of the government and the lack of research on energy data management. The highly decentralized system creates scattered data, making no single agency responsible for coordinating the overall management. One recommendation is for the future research to take these factors, as well as other aspects, when studying energy data management in India. Such a qualitative approach is valuable because it pulls the attention to 'missing data', which is often not as neutral in its effects as it seems.

Keywords: energy data, India, data gaps, energy studies, off grid, renewables

---

Acknowledgements

Writing this dissertation has given me the opportunity to interact with many people and enrich my own repertoire of knowledge. First and foremost, I would like to show my sincere gratitude to my supervisor, Dr. Michael LaBelle, who has played an instrumental role in the successful completion of this dissertation. All thanks to his class lectures, which developed my interest in the energy domain! He always knew the right concoction of geniality and severity which steered me towards the right direction. Also, his door was always open for my simplest queries and his rigorous class lectures were immensely helpful in moulding my dissertation.

My heartfelt thanks to all the respondents who took out time from their busy schedule and spoke at length about their experiences. I'd also like to acknowledge Mr. Bansidhar Bandi (NITI Aayog) and Mr. Vijay Menghani (CEA) for connecting me with a most of my interviewees. Their narratives form the backbone of the study and without their cooperation and timely assistance, it would not have seen the light of day. Lastly, I would extend my gratitude to CEU, the Department of Environmental Sciences and Policy and the Alumni Scholarship staff for equipping me academically and intellectually to successful completion of my dissertation and providing me the chance to explore my area of interest.

Last but not the least, I would like to thank my family and friends who have always supported me unconditionally in my endeavours. Mere words would not be able to express my gratitude to Debapriya, who committed as much effort on this work as I did, if not less. Hence, this work is dedicated to you!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CEA	Central Electricity Authority
EDM	Energy Data Management
EIA	Energy Information Administration
GIS	Geographic Information System
GoI	Government of India
MNRE	Ministry of New and Renewable Energy
MoC	Ministry of Coal MoP Ministry of Power
MoPNG	Ministry of Petroleum and Natural Gas
MoSPI	Ministry of Statistics & Programme Implementation
NITI	National Institution for Transforming India
NSSO	National Sample Survey Organization
PPAC	Petroleum Planning and Analysis Cell

1. Introduction

Over the time, data has grown to become one of the most valuable entity, which is central to how knowledge is produced, business conducted, and decision-making enacted (Kitchin and Lauriault 2014). Similarly, energy data, which constitutes all relevant data about energy, is central to formation and analysis of energy policies, empowering new business models and fighting climate change (Liu, et al. 2017). Although there have been significant developments in the ways data is produced and utilised, there are still several challenges encountered in this area.

One of the challenges is the availability of data and in developing countries like India, it often acts as a stumbling block in research and policy cycle (Rai, Tongia, et al. 2017). The non-existence of a central entity in charge of collecting and disseminating energy data renders the data scattered and poses scarcity of data which, thus, affects the efforts put for effective modelling, research and analysis (Rai, Tongia, et al. 2017). In some cases, it also limits effective public participation and NGOs often have to rely on assumptions to process the data for energy policy making and analysis. In countries like India, which are facing rapid expansion in terms of urbanisation leading to more people switching to modern energy alternatives, the energy policy decisions bring in the opportunity to leapfrogging (Tongia, Rai and Shrimali 2017). With the introduction of newer projects like smart grid mission, smart cities and solar mission, another issue is that the efficacy of these is less understood without relevant open data for the analysts. This lack of relevant data is often criticized by researchers because it acts as a bottleneck in the system.

Studies around energy data are predominantly about the economics and performance of schemes. These studies are typically built on a combination of technical parameters and

finances, in other words, quantitative data (Lovell and Watson 2019). Scholars focusing on data issues within the energy sector are limited (Lovell and Watson 2019). This does not deliver the full picture of energy data management and hence, the research aims to fill the gap about the fundamentals of energy data management system in India and the challenges it face. The aim of the research is to highlight, through a qualitative study, factors aside from infrastructural and technical, which affect and pose challenge to the development of energy data management system existing in the India. By studying this field, I argue that there are four major challenges – the existing collection mechanism itself, present data gaps, socio-political motives of the government and lack of research in the data management field.

1.1 Aims and Objectives of the Research

The aim of the research is to investigate the challenges which are faced by the existing energy data management system in India. The research attempts to answer the following research questions:

1. What are the challenges faced by different stakeholders involved in the existing energy data management system in India?

In doing so, the research project primarily tries to examine the importance of energy-related data and how its private and government stakeholders interact with it for decision making in India. This research will fill in the academic knowledge gap of the energy-data dissemination in India and hopefully will motivate more researchers and advocates to work on this area.

To achieve these aims, I conducted field research including interviews with different stakeholders who are experts in the field. As an outcome of interviews with experts, the research tries to tease out challenges faced by the relevant stakeholders in the process of data uptake for analysis and provide recommendations. A more detailed description of the methodology and the research design has been discussed in the methodology section.

1.2 Outline of the thesis

The following chapter of the thesis is the literature review. This chapter will review the literature available on different aspects of energy data management in India, under its subsections. The first will define and introduce the concept of energy data management and the terms associated with it, in addition to the major energy data management institutes around the world. The later subsections discuss the developments that have happened over the time and current governing institutes with a brief summary of legal mandates which govern the collection of energy data in the country. The last sections of the chapter attempt to review the importance of energy data and the issues reported around it.

The third chapter discusses the methods used in the research and the literature around these methods. During the chapter, the highlight will be on how the data was collected using interviews and the way it was analysed. Chapter four will introduce the results and discussions of the data collected from the interviews. This chapter also fills in the gap which we find in the literature reviewed in attempt to answer the research questions posed in the beginning of study.

Lastly, chapter five takes all the information and draws important conclusions about energy data management in India. In doing so, we provide some recommendations based on the study results.

2. Literature Review

The purpose of the chapter is to furnish a comprehensive understanding on the topic of research – 'Challenges to energy data management in India' and to establish the gaps in knowledge, hence justifying this research study. In attempting to define energy data, this chapter begins with introducing energy studies and data studies. It is followed by a rigorous and extensive review of the existing scholarship in the area of energy data, the past and current institutions in existence and the importance of energy data management. It is important to introduce the definitions and scoping that this current research adopts as this represents how the collected information is analyzed and understood. During the entire chapter, several examples and narratives have been drawn upon to represent different views in the field. The study is focused on India, but the author has taken the liberty to include relevant studies and examples from outside to provide substantive description of the current state of the art in the broader field and also, to provide a foreground of the current study.

The current chapter begins with introduction to energy data and defining key concepts and terms in the field. In the following sections, the major developments in the sector as well as the current functions of the institutions, both legal and administrative, are discussed in the setting of India. Lastly, the author overviews the available literature discussing the importance of energy data management followed by the critiques and issues reported by the scholars and experts in the field.

The aim of the chapter is to highlight the research gap that this research project contributes to. The focus will be on critically analyzing overall benefits of the energy data as well as the issues faced in proper management of it. Since the country specific literature on such an essential issue is very limited, the current research would solely focus on analyzing the challenges

discussed by different stakeholders in the sector. Some external factors might be important in the study because they have an impact on the relationship between the coordinating bodies and therefore, the development of the current system. Hence, some of the further factors such as history, perception, politics and so on have not been taken into consideration which can affect the study.

2.1 Definitions and introduction to energy data management

A report on energy data gaps by Indo-US energy dialogue described energy data in an expansive way as "*all the data that has any link to the energy sector*" (Prayas Energy Group 2015). Energy data does not have a universal definition agreed upon, but rather, can be loosely defined as any form of data or information related to energy – generally obtained from the processes like generation, transmission and consumption. Since the term has ambiguities surrounding it, it can incorporate an array of information related to energy, ranging from the potential solar irradiation in California to the consumption pattern in public buildings. Even the data about the finances can be sometimes clubbed under this definition. In the following text, some of these terms have been explained.

Similar, the term 'energy data management' has not been a recipient of exacting definition but can be understood simply as dealing with energy data in a way which eases its utilization for all kinds of purpose. In a report formulated by Dukkipati and colleagues (Dukkipati, Iyer and Sreenivas 2014), energy data management (EDM) systems have been collectively defined as "*institutional mechanisms and processes to collect, store, validate and disseminate data in a timely manner*". In other words, energy data should be organized such that it could be utilized for analyses going beyond usual economic projections and performance calculations.

If one looks beyond the energy sector, a systematic and thorough study of data called 'critical data studies' has been generating relevance within the larger body of scholarship (Dalton and

Thatcher 2014). The approaches in critical data studies are more qualitative and stems from the fundamentals of critical theory and data science (Kitchin and Lauriault 2014). Although this method has brought the role of social sciences in a new scientific field like big data, it is possible to expand its approaches to energy sector particularly for energy data studies. Critical data studies can be a significant component to understand energy data studies because it can bring social science perspective to energy studies and that too when big data has become so significant. The first such study in the energy field was published recently by (Lovell and Watson 2019) in order to introduce critical data studies as a social sciences analysis for scarce data (as a contrast to big data) in the energy sector. One of the core aim of the study was to position their research as a contribution to big data studies, but in contrast, as they dealt with scarce data (Lovell and Watson 2019).

With the above description of energy data management, it is clear that a good energy data management mechanism would satisfy the needs of a wide variety of data users and would be easily accessible for such users. The data obtained from such a system can be used for a variety of functions like financial calculation, peak prediction, environment monitoring and so on. When such data is not available at a desired quality or easily accessible to its users, it is referred to as a 'data gap' (Prayas Energy Group 2015).

The most common forms of energy data include the data from the generation and the consumption side. Generation here refers to the production of energy which is also called supply-side or upstream (OECD 2019). Consumption denotes the end-use of the generated energy and terms like demand side or downstream are associated with it (OECD 2019). Depending on the scale and location, one can categorize energy generation as centralized or decentralized (EPA 2018). The definitions of centralized and decentralized are usually expanded for electricity production and hence some forms of energy such as biomass, oil and gas are lesser used in this definition. According to (EPA 2018), 'Centralized generation'

facilities produce energy in large-scale and then distribute via transmission lines to end-users located away from the generation source. These transmission lines are typically called grid and end-users connected to the grid are said to be on-grid. Decentralized or distributed generation is vaguely defined as small-scale electricity generation. A study by Pepermans and their colleagues (Pepermans, et al. 2005) compares several definitions and finally conclude to the definition proposed by (Ackermann, Andersson and Lennart 2001) which factors (i) the type of connection and (ii) location of the generation unit instead of generation capacity. When the energy generation takes place in vicinity of end-use, it is usually decentralized. Such facilities are not connected to the grid, and therefore, are said to be off-grid systems or simply off-grid. When compared, the centralized generation is dominant but the smaller distributed generation capacities have always been existed in the form of batteries, biomass and more recently solar photovoltaics. The distribution of energy to the end-user is called transmission and it is facilitated by high-voltage lines in order to reduce losses. Consumption is defined according to the end-use. Majority of the energy in India is consumed by four sectors: agriculture, industry, commercial and residential sector (Prayas Energy Group 2015).

When looking outside India, there are several well-developed energy research institutions working both at global as well as at the national level. The structure of these institutions can be broadly described either as independent unit or as a distributed unit under central entity (Chikkatur and Chakravarty 2008). International Energy Agency or IEA was formed in the wake of oil embargo with the charge to promote member country energy security – being autonomous agency of OECD, but now have four areas of focus (IEA 2019). The member countries of the IEA are obliged to submit questionnaires to them while it is voluntary for non-members. Important information about some of the key institutions from countries such as USA, Australia, China and so on have been summarized in the table below.

Table 1. Institutions for energy research in other countries (adopted from (Chikkatur and Chakravarty 2008))

Name	**Country**	**Host institution**	**Major activities**	**Accessibility of information**
International Energy Agency (IEA)	OECD+ (OECD and other countries)	Autonomous agency under the OECD	- Energy policy advisor - Collection of energy data - Analyses of energy data (modelling, forecasts etc.) - Publishing reports - Development of data management systems in OECD+	Most reports available on web while recent data is limited to paying subscribers
Energy Information Administration (EIA)	United States of America	US department of energy (DOE)	- Collection of energy data, developing surveys - Policymaking, analyses - Modeling - Publication of routine reports	Subject to US Freedom of Information Act
Australian Energy Statistics	Australia	Department of the Environment and Energy, Australia	- Collection and statistical analyses of energy data - Routine report with dataset - Collaborate with private and research sectors	Non sensitive data is open access
Energy Research Institute (ERI)	China	Chinese National Development and Reform Commission (NDRC)	- Monitoring - Policy and technical advisories - Publishing routine journals and reports	Reports available on website, but payment required for full-text

The above table is a list of institutions dealing with energy policy research in different countries, their major activities and how much information is made accessible by them. While some of them provide datasets and open access reports on their website, some require payment for disseminating the useful data.

2.2 Developments in energy data management

The section attempts to recognize the developments in energy data management in India by describing the power sector and the structures governing it. The rationale behind writing this section is to introduce and show the developments in energy data management over the years. The developments in the sector have to be in a form that makes the entire sector under one umbrella. There are numerous users of energy data – researchers, government, consumers and so on. Hence, it is important for the mechanisms which regulate the energy data to be consistent with energy sector bodies and practices of the country (Dukkipati, Iyer and Sreenivas 2014).

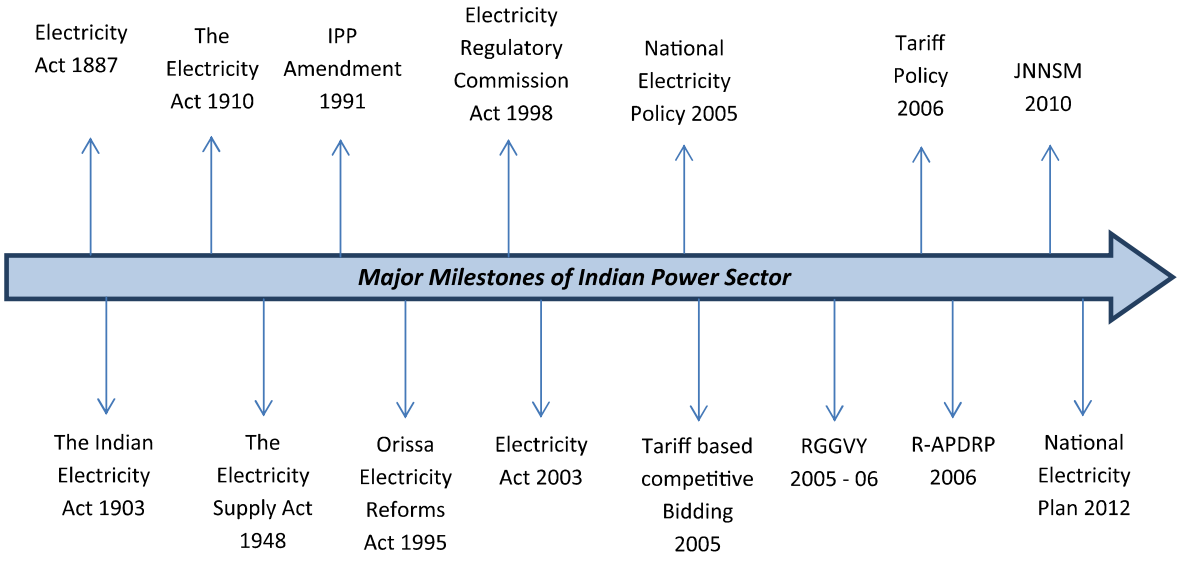

Figure 1. Evolution of Indian power sector (Mukherjee, Dhingra and Sengupta 2017)

Prior to the independence in 1947, there was not any law for regulating the data collected from energy sector, but some energy surveys were undertaken (Chikkatur and Chakravarty 2008). The electricity generation and supply were mainly in the urban areas and regulated by the private entities (Mukherjee, Dhingra and Sengupta 2017). Hence, the first legislation to regulate the electricity sector was introduced in the 18th century and called as the Indian Electricity Act (1887), which got further modified in 1903 and then in 1910 (Mukherjee, Dhingra and Sengupta 2017). Post-independence, the Electricity Supply Act (1948) led to the

development of separate "electricity boards" for states (SEBs) along with a national entity named "Central Electricity Authority" or CEA, which is discussed in detail in the sections below (Mukherjee, Dhingra and Sengupta 2017). In the next years, the vote bank politics led to huge debts in the system (Joseph 2009) leading to the introduction of private industries in the sector. To determine the tariffs at both central and state level, Electricity Regulatory Act or ERC Act 1988 was introduced (Mukherjee, Dhingra and Sengupta 2017). The much welcomed and revolutionary Electricity Act 2003 which led to the creation of CEA has been discussed in the detail in the section below.

Census Act, 1948 came into effect to formalize the national scheme as well as duties of the census officials. Till 1948, the census in India was conducted on an ad-hoc basis, however, the regulations have now made it mandatory in every ten years. Although every citizen is bound to participate and provide best knowledge for the census questionnaires, it is up to the Census Commissioner to decide on which information from the Census data should be released (Dukkipati, Iyer and Sreenivas 2014).

2.2.2 Central Electricity Authority (CEA)

In the year 2003, the government passed The Electricity Act (2003) which aimed to improve transparency in the electricity sector of the country making Central Electricity Authority (CEA) responsible for the gathering and distribution of electricity data (Ministry of Law and Justice 2003). The act also granted CEA the right to collect data from any licensed power generator and hence became a major source of data in India (Ministry of Law and Justice 2003). In addition to any licensed power generator, CEA is conferred the authority to collect data in any entity involved in electricity transmission, distribution and trade as well (Dukkipati, Iyer and Sreenivas 2014, 32). It is important to note that the data has to be provided to CEA in the format specified by CEA and it could vary from case to case.

Currently there are six wings under the CEA which are responsible for planning, thermal, hydro, grid operation and distribution, economic and commercial and power system (Dukkipati, Iyer and Sreenivas 2014, 36). It was created under the Electricity (Supply) Act (1948) and is currently responsible for the collection and distribution of energy-related data with daily, monthly as well as annual reports. According to Section 73 of this act, CEA is also responsible to make the information as well as the investigations available for public use. In the table below, one can find a non-exhaustive list of CEA reports and their frequency of publication.

Table 2. A non-exhaustive list of CEA reports (adopted from (Dukkipati, Iyer and Sreenivas 2014))

Type of data published	**Frequency of publication**
Unit-wise Generation and Outages	Daily for thermal, nuclear and hydro across all sectors (central, state, private)
Coal Stock Position at Thermal Power Plants	Daily for imported as well as domestic coal
Hydro Reservoir Level	Daily
Installed Capacity	Monthly for thermal, nuclear, hydro and renewable
Capacity Addition Status/Review	Monthly for thermal power plants
Transmission line Length and Substation capacity	Monthly aggregated at voltage-wise sector-wise
Unit-wise Generation and PLF with year-on-year comparison	Monthly for thermal, nuclear and hydro across all sectors (central, state, private) and fuel –wise
Power supply position	Monthly for energy as well as peak requirement
Captive Generation	Annually
Outage Plan for next year	Annually
Performance of thermal power stations	Annually

Performance of hydro power stations and hydro power potential	Annually
Baseline CO2 database for CDM consisting of all generating units, net generation during the last year and absolute CO2 emissions	Annual or as required

The Electricity Act (2003) brought a transparent framework for the structuring of power sector and made metering compulsory. Furthermore, the act made the provisions relating to electricity theft strict and allowed the establishment of off-grid systems for generation and distribution in rural and remote areas (Pal 2013).

2.2.3 NITI Aayog (formerly Planning Commission)

NITI Aayog is responsible for formulating the national energy policy, to develop the India Energy Security Scenarios (IESS), exploring data management opportunities at the national level and collaboration with the IEA and EIA (Iyer, Kumar and Mathew 2016).

Modeling plays an important role in decision making for any sector. It was in the year 1963 that the first comprehensive survey of energy resources and projections of future energy demand was undertaken which collected, organized, made demand projections and provided policy recommendations for further planning (Sankar 1985). Initial modelling practices were carried out with the support from foreign sources such as with the US and UK, which however, proved to be 'ineffective' because they were not truly able to examine the Indian conditions for the models (Chikkatur and Chakravarty 2008). Although the initial modeling endeavors were short-term, rigorous practices came into effect with coordinated efforts to fit the Planning Commission's national economic growth plans. Modeling practices such as demand predictions have been coordinated and carried out by the Working Group on Energy Policy (WGEP) set up at 'Planning Commission' at the national level. Planning Commission of India has been the focal point of policy recommendations within the government (Chikkatur and

Chakravarty 2008) which was however, dissolved into another institute 'NITI Aayog' in 2014 when the new government took the responsibility. Presently, NITI Aayog takes lead in the energy data management efforts of the country by coordinating between relevant ministries, private organizations and research institutes. Based on the inputs of different stakeholders and workshops, NITI Aayog is leading the "*development of a roadmap for improving energy data management in India*" (NITI Aayog 2019).

2.3 Energy data management in India - Journey of energy data

This section of the current chapter brings forward the complexities of the energy data management in India and in addition to it, describe the legal and institutional structures which shape the foundation of energy data management in the Indian context. The sections, discussing data collection and dissemination, also review the instruments and mediums used for these processes.

2.3.1 Institutions and governance

The constitution of India has classified each sector to be administered by either government at the center (national) level or government at the state level or both. For example, central government is responsible for administering sectors on the supply side such as oil, coal and gas. However, electricity is a concurrent subject which means it is dealt at both central and state level. In addition to it, sectors on the demand side such as rail transport and aviation transport are administered by center while agriculture is a state subject. The figure below summarizes major institutions at different levels – central, state, research institutes and even non-profit organizations that collect data.

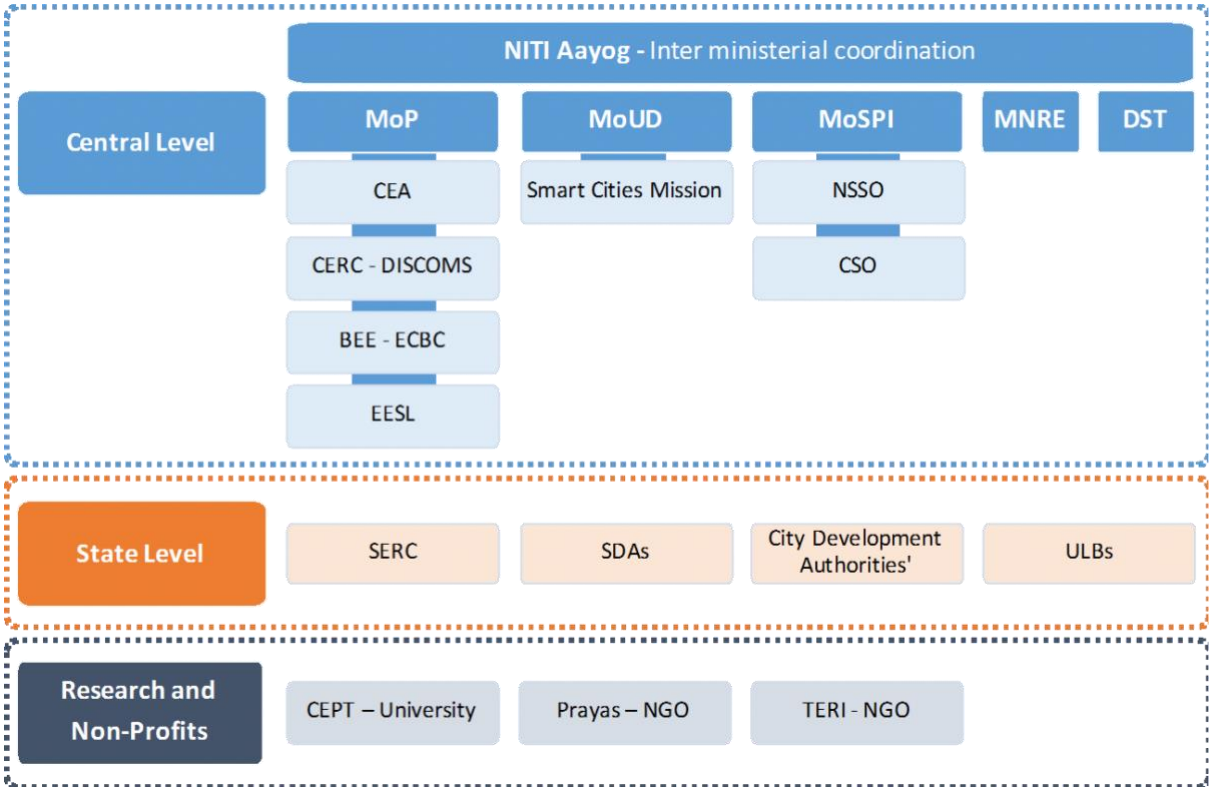

Figure 2. Government and Non-government Institutions in India involved in the building stock energy data framework (Iyer, Kumar, et al. 2016)

NITI Aayog, which is discussed in the previous section, coordinates with the different ministries and other non-governmental stakeholders. The Ministry of Power (MoP) heads CEA and BEE and has specific mandates to collect, analyze and distribute the data. Department of Science and Technology (DST) is responsible for maintaining the online data portal data.gov.in. Although the figure above is for building stock energy data collection, it provides a coherent visual of the structure. In addition to the government ministries, there are numerous stakeholders which benefit from the energy data in different ways.

Majority of the administration in the energy sector in India is held across five departments or ministries under the central government (Iyer, Kumar, et al. 2016). These are Ministry of Power (MoP), Ministry of Coal (MoC), Ministry of New and Renewable Energy (MNRE), Ministry of Petroleum and Natural Gas (MoPNG) and Department of Atomic Energy (DAE). Prior to the formulation of these five ministries, India had combined ministry called The Ministry of

Energy Sources. This was dissolved in 1992 which paved the path to the formation of independent ministries heading distinct departments with specialized interests (Ministry of Power 2019). In addition to these ministries, there are ministries for energy demand sectors like agriculture, transport and industries at both central and state level. Currently, the data is collected by nodal agencies or through state.

Table 3. Legal system and developments (adopted from (Dukkipati, Iyer and Sreenivas 2014, 7-10))

Type	**Year**	**Legal act**	**Main provisions**
Fuel	1948	Oilfields Regulation and Development Act	Oilfields act provides central government the power to create rules for mineral oil mine owners/lessees. Hence, these rules can make it mandatory for the mine owners/lessees to submit periodical reports on their operations. **Petroleum and Natural Gas Rules, 1959 (amended in 2006)** were created under this act making it compulsory for the agency to submit *"geological, geophysical, geochemical.., reports, analyses…prepared in respect of petroleum operations"* at the earliest and without any cost.
Census	1948 (amended in 1994)	Census Act	The act in original formalized the administering of Census with questionnaires devised by the Census Commissioner. With the act in practice, it becomes necessary for all citizens to participate in the surveys under the Census Act.
Fuel	1955	Essential Commodities Act	In the Essential commodities act, the central government is granted the responsibility to collect any information or statistics in order to ensure public interest. The data obtained can be used to regulate or prohibit functioning of essential commodities which could potentially be unfavorable for public interest. In addition to typical supplies, the essential commodities include coal (including derivates like coke etc.) and petroleum (and other derivatives). Later in year **1999, Petroleum Products (Maintenance of Production, Storage and Supply) Order** came into effect under the Essential Commodities act. The order mandates all the oil refining companies to submit information on their produce (including derivatives) on a per-month basis and the planned production programme for the next month to the central government.

Fuel	1957	Mines and Minerals (Development and Regulation) Act	The Mines and minerals act provide government (both central and respective state entities) the authority to survey any mine facilities and inspect its records. With this act, the government entity can sanction any personnel an entry inside the mine/facility and inspect records or survey any person who's in-charge of the facility.
Fuel	1974 (amended in 2011)	Coal Mines (Conservation and Development) Act	Under this act, every owner or agent is bound to make available data on generation and transfer of coal, washery and process products. In addition to these, information regarding the operation conditions as well as opening/closure of mines should also be furnished as demanded.
Energy Conservation	2001	Energy Conservation Act	Consumer of energy has the duty to make available all the information and data related to energy consumed, and this act granted Bureau of Energy Efficiency (BEE) the mandate to exercise this legal statute.
Electricity	2003	Electricity Act	Electricity act empowered the Central Electricity Authority (CEA) to gather any kind electricity data from all the entities/corporations involved in electricity sector. CEA is then made responsible for disseminating all the data to public domain as well.
Fuel	2006	Petroleum and Natural Gas Regulatory Board Act	The act formalized data gathering by making the regulator responsible for maintaining a data bank of its downstream oil-gas activities. In addition to the duty of data collection, the regulator has been granted the right to verify the data provided by downstream entities at the same time maintaining confidentiality of the information.
Statistics	2008	Collection of Statistics Act	The act came into effect to empower the central and state governments in gathering data for statistics on any subject. It also mandated such agencies to provide data in public domain and the failure to provide data by any stakeholder a punishable offence. In 2011, the Ministry of Statistics and Programme Implementation (MoSPI) informed a new rule called **Collection of Statistics Rule, 2011** which mandated appointment of a nodal officer by the Central Government accountable for managing and improving data collection and records.

The structure of the energy data flow for the conventional electricity sector in India is governed by Ministry of Power (MoP) at the center (national) level along with the respective state energy

departments. Ministry of Power depends upon the Central Electricity Authority (CEA) for the technical guidance and policy analysis.

2.3.2 Data collection in India

Authors (Iyer, Kumar and Mathew 2016) tellingly charts out the different ways in which the data collection takes places in India. The data from energy generation at first is usually collected by the stakeholder, depending on the ownership of the generating unit. The data is then transferred to the responsible state/central government bodies in the specified format. Since electricity is a concurrent subject, handling the data is responsibility of both the state and the central side. As discussed in the previous sections, CEA is the main body responsible for collecting the data related to electricity. The figure below summarizes the CEA's involvement with institutions collecting energy data. It can be seen that State Load Dispatch Centers (SLDCs), Regional Load Dispatch Centers (RLDCs) and Electricity Regulatory Commissions (ERCs) report data to CEA.

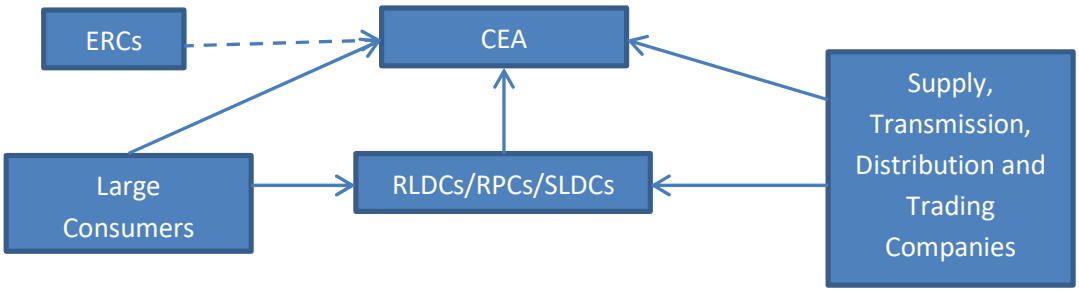

Figure 3. Data collection by CEA (Dukkipati, Iyer and Sreenivas 2014)

In a similar fashion, the ministry of new and renewable energy (MNRE) also collects data from state nodal agencies, NGOs and private organizations.

There are two major bodies given the duty for collecting consumption side energy data – National Sample Survey Organization (NSSO) and other separate ministerial bodies like for agriculture, transportation and so on.

Collection of energy data from the generating sites is usually controlled by the automated metering systems and transferred to the authorities via internet. For example, CEA has implemented an Information Management System (IMS) to gather data from different licensees. However, the data collection for the end-use usually takes advantage of mixed tools including conventional tools like surveys and modern tools like electricity metering systems. At locations where metering systems aren't automated, responsible electricity boards collect the data from the households manually through physical surveying. It is to be noted that there is no mandate for collection of data from units generating power less than 1 MW.

2.3.3 Data dissemination and usage

The data collected by the CEA is published in the form of reports on its website on a daily, month and annual basis and covers information like unit-wise generation, reservoir levels for hydro, installed capacity and so on. The annual report by CEA called 'All India Electricity Statistics' contain accumulated data from the daily and monthly reports in addition to other data such as system losses and captive generation for plants generating power more than 1 MW. This annual report is generally published with a one to two-year long delay and only in printed form. Apart from publication of datasets, CEA also conducts research within the department and publishes long-term demand forecasts in an interval of five years in order to overlap with the Indian planning cycle (CEA 2013). In addition to CEA, the state and regional dispatch centers also disseminate important data, typically on their website. However, when checked online on their websites personally, it was noticed that the quality of the data varies from one center to other.

Similarly, MNRE also publishes the aggregate data on its website. As far as energy statistics is involved, Ministry of Statistics and Program Implementation (MoSPI) is responsible for the official statistics release in the country and also for the energy balance. In addition to it, separate ministries and departments disseminate their own data and publications routinely. For example, in the transport sector, different ministries publish statistical summaries over the annual basis in their reports. The yearbook by Ministry of Railways contains data such as fuel consumption as well as costs incurred.

In the recent years, both the government and non-government institutions have put in worthy efforts for managing energy data. Rai and colleagues (Rai, et al. 2017) mention and critique various initiatives and projects that have been introduced to integrate and disseminate energy data in the public domain. A website called *data.gov.in* was created by the Government of India in the year 2012 to disseminate data on a single online platform in a cohesive manner from various ministries. In addition to the fact that this website collects already established data, (Rai, et al. 2017) existing study has critiqued the effectiveness of this initiative. Several other initiatives in different forms are being undertaken, for example, the annual report "Energy Statistics" published by the Ministry of Statistics and Programme implementation (MOPSI) which complies data from diverse sectors like natural gas etc.

State electricity regulatory commissions have played an important role in improving the quality of the energy data, such as in Maharashtra state of India (Prayas Energy Group 2015). The utilities in the state have been sharing rich hourly data on electricity usage and cost, although in PDF format till recently (Rai, et al. 2017). Data portals are another form of channels which can utilize bundles of information and use communication technology.

2.4 Importance and usage of energy data

Data in every form is quite telling of the intervention, and this section of the chapter discusses the significance of energy data and its role in the niche and overall system. Data is one of the primary parameters that drives vital necessities of our daily lives. For example, heavy volumes of data is generated every day and routinely gathered by governmental and non-governmental institutes for fulfilling multiple purposes – forecast, tax, investment, health, tariffs and so on (Kitchin and Lauriault 2014). The work by Professor Shaffer (Shaffer 2017) for CIA called 'A Guide to the Application of Energy Data for Intelligence Analysis' introduced the readers to numerous usage of energy data in the world.

2.4.1 Effectiveness of policy

In a study called 'Effective energy data management for low-carbon growth planning' by Liu and colleagues (Liu, et al. 2017), presence of high quality energy data has been linked to effective analysis and policymaking for the environment. The study examines data energy data management in four countries – Germany, Canada, UK and US and presents a framework which can be used for evaluation of national EDM systems (Liu, et al. 2017). It is well understood that the data acts as an instrument to measure the effectiveness of the policies implemented and even for the formulation of the policies (Rai, et al. 2017). In the energy sector, the formulation of effective policies rely on analysis of data and availability of accurate and accessible data smoothens the entire policy cycle process (Dukkipati, Iyer and Sreenivas 2014).

2.4.2 Reporting and monitoring

The country's obligation under its Intended Nationally Determined Contributions (INDSs) submitted at the COP21 Paris lists MRV or Measuring-Reporting-Verification activities as a top priority (Rai, et al. 2017). The projects aiming to reduce energy poverty, sometimes also reducing pollution, hinges on data obtained from public domain for their operations such as researching, reporting, policy making and planning. Numerous energy sector programs and

schemes have been implemented in India by the government and their performance can only be evaluated if the proper data is available.

2.4.3 Modelling and projections

Dealing with numerous energy challenges in the future requires development of a extensive capacity in energy and economy modelling analysis and forecasting using complex modelling tools (Chikkatur and Chakravarty 2008). Regular collection of large quantity of data is the first and most important step which when followed by statistical analysis allows the energy modelers to get a better understanding of the energy system and its determinants altogether (Chikkatur and Chakravarty 2008). In addition to create a better understanding of the system, these modeling practices also provide vital feedback to inculcate improvements in the data collection process of the system.

2.4.4 Business and innovation

The trends of electricity consumption often correlate with the economic growth pattern, with similar trend being followed in the employment rate as well. For example, the case of China where the change in electricity consumption rates directed the analysts to correctly predict the 2015 downward trend in economic growth well in advance (Shaffer 2017).

The data from energy also plays an important role in business strategy formulation and in the development of innovations. Innovation and creation of new business models in the energy domain would also be accelerated when more data is made available. Such was the case with California Solar Initiative (CSI) which enabled the businesses to utilize data and adapt rapidly to meeting demands of the market by introducing new plans and making good use of subsidies (Rai, et al. 2017). In the U.S. and around the world, several start-ups have been initiated whose entire model is based on unlocking the potential of energy data from households.

2.4.5 Energy standards

A telling example from India about the need for proper dissemination of energy data for policy making is the process of improving the Minimum Energy Performance Standards (MEPS) for key electronic appliances (Phadke 2014). The process is conducted with utmost detailed study of the entire supply chain including technical feasibility and incremental cost for both consumers as well as producer. These analyses are reported in detail for the public domain. In India, although the revisions for MEPS have started to take place, the comprehensive assessments aren't publicly available (Rai, et al. 2017). As a result, the MEPS for electrical appliances in India are among the lowest in the world (Rai, et al. 2017).

Monitoring the performance of the building stock is another important usage of the energy consumption data. a report about energy datasets for buildings, (Iyer, et al. 2016) establishes that the presence of robust dataset about the building's energy usage enables implementation of energy efficient design and overall management in addition to enhanced procurement of end-use equipment.

2.5 Critiques and opportunities

There have been several critiques and issues which have been reported about the energy data in general as well as in India. This section discusses some of them in addition to the suggestions and feedback for the improvement of energy data management in the country.

2.5.1 Data collection and usage

The first step in energy data management - collection of data, is done through a variety of tools such as automatic metering, surveys, census and so on (Dukkipati, Iyer and Sreenivas 2014). As far as electricity is concerned, CEA has the mandate to collect data from any related entity in any form. In addition to CEA, the state and central regulatory authorities collect important data however they are not mandated to share it with public (Prayas Energy Group 2015, 17). The issue becomes prominent in the rural areas where poor supply quality and insufficient metering infrastructure persists. Several schemes for rural electrification like *Deen Dayal Upadhyay Gram Jyoti Yojana (DDUGJY)* and *Restructured Accelerated Power Development and Reforms Programme (R-APDRP)* have received sufficient funding, however, the data collection and distribution is still an issue (Prayas Energy Group 2015, 17).

Various regulatory bodies collect important consumption data which doesn't make it to a centralized body, hence the data remains scattered across them making it difficult for users to get hold of it. These include sectoral ministries like department of agriculture, railways and so on. National Sample Survey Organization or NSSO is an organization specialized and responsible for undertaking the consumption pattern surveys in the country. Although surveys are conducted by the National Sample Survey to collect data, their energy consumption data was found to be insufficient for policy making (Rai, et al. 2017). Similar energy data inconsistencies in the European region were also reported by (Thollander, et al. 2015) on small and medium sized industries. (Thollander, et al. 2015) advocated for a consistent structure and

taxonomy to be developed and followed in order to make available more 'textured' data for the users, specially governance structures.

In India, collection of energy data from electricity generating systems with capacity less than 1 MW is another issue because there is no mandate for the users to report it. Hence, there is almost no data available on such small and decentralized power generating units. Majority of such units are solar rooftop because of the policy push by the government for renewable energy technologies even at household level. Issues in collection of off-grid data has been discussed in detail in the section following.

It is not just the collection of data about the end-use consumption but also the statistical studies which require improvement. (Christensen and Hain 2017) point out several types of challenges faced by them in realizing a statistical overview of investments in the energy sector, particularly renewable energy sector. Their study about investment on renewables conclude that there is still a lack of datasets to capture the sector wholly and hence several areas in statistical analyses as well as availability of data requires improvement (Christensen and Hain 2017).

Since solar is placed at the forefront of renewable energy transition, new advancements in the policies and investments occur frequently. The data issues are widespread in this case too, with new data being generated more often and the stakeholders demanding new data for research. In case of India's solar sector, there is an "all-round paucity of data" and the quality has been questioned too (Rai, et al. 2017). In the case of non-availability of overall data from the government, researchers rely on privately owned data from commercial sources which often suffer from validity issue (Rai, et al. 2017). Even Ministry of New and Renewable Energy (MNRE), the responsible ministry, does not generate or disseminate data on its own but rely on other central and state agencies for collection of data.

In contrast to the data issues from solar energy sector in India which were mentioned in the section above, the accessibility of such data from United States is admirable. The data from solar projects even from the project-level is made available to the public domain (Rai, et al. 2017). As a result of making solar data open, a lot of extensive studies have been already carried out ranging from new business models, modeling, social and behavioral aspects. In particular, the open data from California Solar Initiative (CSI) has proven to be effective in enhancing cost-effectiveness of solar programs and their subsidies (Rai, et al. 2017).

2.5.2 Creation of centralized entity

A more recent perspectives paper, (Rai, et al. 2017) discusses the problems India faces in handling energy data and attributes it to lack of "*a singular central body in charge of maintaining and dissemination India's energy data, let alone analyzing it*". Although it draws upon example of U.S. EIA in the commentary, the study does not base it as the right institutional context for India. Similarly, even the institutions in China, Brazil and South Africa have their own differences and cannot be used to draw similarity in this context (Rai, et al. 2017). Along with the supply of robust datasets, such an entity would improve the understanding of newer introduced projects by the government. However, to leverage such benefits would demand careful planning of policy tools and rightful implementation with "course-correction as learning accrues over time" (Rai, et al. 2017).

In one of the correspondences of Nature journal, (Tongia, Rai and Shrimali 2017) argue about establishing a central body for dealing with energy data in the country. The absence and quality of useful data-sets becomes an obstacle in conducting the research as well as undermines policy and regulatory compliances. Often, data is collected from different entities and then reported as gross averages and lack timeliness. The methodologies and assumptions used by these different entities also pose a challenge to the correct utilization of energy data-sets obtained. (Tongia, Rai and Shrimali 2017) propose the creation of a central Energy Information Agency

to co-ordinate with the existing entities managing data, such as Central Electricity Authority (CEA), Petroleum Planning and Analysis Cell, and with the state and private industries.

In addition to the research and policy making, such an entity would help fulfil the commitments under the Paris climate deal which also include development of mechanisms for measuring, reporting and verification (MRV). Strengthening MRV is an important commitment to comply with the enhanced transparency mechanism established under the Paris Agreement. (Rai, et al. 2017) mention the possibility of leapfrogging other such EIAs, for instance, by introducing quality data with more granularity utilizing the feeder meters and smart grids proposed under National Power Portal and other new schemes (Rai, et al. 2017).

2.5.3 Off-grid energy data management

Though the centralized generation of energy or on-grid system has been dominant over the times, the growth of smaller decentralized capacities evidently suggests that it is capable of overcoming shortfalls of the centralized generation architecture. For example, Denmark has vastly reduced its dependency on the imported fuels over the last three decades and the country uses wind energy to generate more than 30% of their national electricity (2003) (Sovacool and Tambo 2016).

The growth of small-scale residential renewables or the energy generated by these 'prosumers' has started to change the dynamics electricity sector in various countries (Lovell and Watson 2019). The dynamics here refer to the change in proportion of consumers buying electricity from the grid, for example in Australia where the demand of electricity uptake from the grid has been on a decrease (Lovell and Watson 2019). Solar has been seen as one of the major factor to this decreasing trend because of the increasing number of residents opting for installing solar photovoltaics (PV) at their home in addition to improvements in energy

efficiency (Lovell and Watson 2019). This has been taken over by the research of (Laws, et al. 2017) on the situation where consumers start detaching from the central power grid.

A qualitative study by (Lovell and Watson 2019) in Tasmania revealed that the users switched to off-grid because of three main reasons – off grid was financially and operationally viable, users felt it was their environmental and social responsibility and lastly, the sense of control led to a feeling of empowerment. The mass disconnection from the grid would eventually result in increase of electricity costs for households which cannot go off the grid for some reason - be it lack of space or finance (Laws, et al. 2017). The situation becomes interesting in the case of Australia where the information about consumers opting of off-grid sources has been historically ignored due to traditional framing of on-grid data usage (Lovell and Watson 2019).

Similarly, in India, there is a lack of regulation to track off-grid systems. The only available data about off-grid systems are from the subsidized schemes, which too lack consistency. Studies about off-grid energy has been predominantly about rural electrification and its feasibility from the economics and business point of view. (Singh 2016) in his paper discusses business innovation and diffusion of off-grid technology options in Indian markets, while (Spangher 2019) conducted a case study in southern India comparing five types of ownership and finance model of off-grid system. In other areas of off-grid systems research, modelling of future scenarios has been conducted but using estimation rather than using actual data.

The lack of data, as stated in the recent research by (Lovell and Watson 2019), has resulted in the issues related to off-grid not given significance and often being ignored from energy policy frame. There is little or no monitoring of off-grid systems in India, making it uncertain as to how many of them are even operating (Dukkipati, Iyer and Sreenivas 2014, 14). This finding is on similar lines with the finding by (Huges 1983) about the impacts of absence of data in the domain. (Huges 1983) notes that the lack of data makes the issue less ‘visible’ and hence not

discussed by policymakers. With regards to the energy sector, the situation results in policy formation which is skewed towards the certain 'existing energy institutions, technologies and cultures' and create hurdles in the development path of radical innovation (Huges 1983).

2.6 Conclusion

India has multiple agencies which are responsible for collecting different parts of energy data. The above sections review the operation of India's energy data management. It is clear that the country has enacted several legal mandates over the years to evolve the system where it stands now and also to empower the government agencies. Although government agencies have the provision to collect various kinds of data, it becomes a major issue when such powers are used only to fulfil the administrative purposes. The section on importance of energy data discusses the multifaceted benefits of well-developed data management in the country. However, it is a matter of concern when the true potential of said structure is not reached.

Since the number of scholars studying data issues within the energy sector is very limited and the primary areas of research in the energy sector are technical and quantitative, it becomes essential to focus on the areas like energy data studies which are less understood and are of utmost importance. The current study attempts to address this gap.

3. Methodology

The purpose of this chapter is to discuss the study's methodological approach from conception to implementation. In addition to further collection and analysis of existing literature on energy data management in India, this study has adopted a qualitative study design. This chapter begins with introducing the readers to the conception of the research and then the setting wherein the study was conducted. Following that, it discusses the process by which the data was collected and analyzed and ends with the limitations of the study.

3.1 Conception of the Research

The initial idea of the research was to study the structure of management of energy data in India and come up with the recommendations at both policy and structural levels. In order to suggest such modifications, a case-study comparing India with countries having a well-developed energy sector was designed in the beginning. Another idea was to discover and conduct a similar case-study on successful management of any other form of data in India and using similar comparative study to provide recommendations for the energy sector in India. This was an interesting idea because the two sectors – energy and other chosen sector, being in the same country setting, could be compared without factoring in drastic geographical differences for the research.

However, it was realized in the beginning of literature review that a thorough comparative analysis would require more than document and policy analysis. To develop a full image of the energy sector, it was decided that historical as well as institutional setting would play a vital role and should be included in the overall design in addition to interviews. Although the above-mentioned idea of comparative study was a novel one, it made sense to concentrate efforts on studying a single sector. Hence, the design as well as aim of the study was modified a bit in

order to focus on the energy sector and particularly on investigating the challenges in data management prevalent in the country through interviews.

3.2 Document Analysis

The purpose of document analysis was mainly to explore the field of energy data and to introduce certain multifaceted concepts. As mentioned in the previous chapter, the scholarship builds around this domain is still at the nascent stage and thus, one of the goals of this current study is to make a small step to contribute and bolster the larger area of research. Researches (Tongia, Rai and Shrimali 2017) have been efficient in asserting the importance of constructing cohesive energy data sets to not only understand the current structure of the field of energy but also to enact policies as per requirement. Significantly, recent research (Kitchin and Lauriault 2014) (Lovell and Watson 2019) has investigated the terrain of critical data study that draws heavily on qualitative design and social sciences to explain varied approaches to energy data. This kind of perspective is believed to bring further nuances to understand energy data studies and also broadens the horizon of the field. Despite the limited researches undertaken, the already existing studies have been monumental in setting the premises of the current study and, simultaneously, attempts to fill in the factual gaps in the literature. In doing so, this study proposes the following question:

> *What are the challenges faced by different stakeholders involved in the existing energy data management system in India?*

3.3 Data Collection, Field Work and Setting

As mentioned above, this study adopted a qualitative design. The selected tool of enquiry was interview because existing researches have proven how the method of interviewing has always been successful in tracing out rich and layered information (ref). Semi-structured questionnaire was used to conduct the interviews which gave the author the liberty to ask relevant open-ended questions without going astray from the main themes of the research. Along with this, special attention was given to maintain anonymity and confidentiality of all the respondents who participated in this study. All the interviews were audio recorded and permission was sought from the respondents to use the excerpts of their narratives for data analysis of the current study. This study received CEU's travel grant to cover the costs incurred in fieldwork.

Ten interviews were conducted over the span of twenty days in the month of May-June. Each interview lasted for about six to ninety minutes and was conducted in English. Prior appointment was fixed with interviewees through face-to-face request, emails and phone calls. When none of these mediums of communication worked, social media was used to connect with the experts. For example, one of the experts leading a government body was contacted through Twitter, while a person leading private solar PV agency was contacted through LinkedIn. Respondents of this study were mainly experts in the field of energy data management. Inducting respondents in the study was challenging because most of the invitation sent out for participating in the research went unanswered and caused inevitable delay. However, the amassed data that eventually unfolded shaped this research in ways unanticipated and uncovered important vantage points. The respondents came from both governmental and non-governmental agencies and this gave the author a wider understanding and better engagement of their perspectives. The interviews took place in their respective offices. The affiliating institutions and the sector have been mentioned in Table no. 1.

Table 4. showing the affiliated institutes and sectors of the interviewees

Serial no. of Interviewees	Representative institute	Sector
1	NITI Aayog	Government think-tank
2	Council on Energy, Environment and Water - CEEW	NGO/Research Institute
3	Council on Energy, Environment and Water - CEEW	NGO/ Research Institute
4	Central Electricity Authority	Government public domain
5	The Energy and Resources Institute - TERI	Academia
6	Lawrence Berkeley National Laboratory	Academia
7	Ashoka Trust for Research in Ecology and the Environment	NGO/Research Institute
8	Ministry of Statistics and Programme Implementation	Government body
9	Off grid private company (name not revealed)	Private Institution
10	Private university in India	Academia

3.4 Data Analysis and Coding

The term 'coding' in social and qualitative research is used to denote a word or a short phrase symbolizing and summarizing a piece of qualitative data (Saldaña 2015). The data could range from verbal and textual material such as interview transcripts to visuals such as photographs and videos. In simpler words, (Charmaz 2001) has explained coding as the "critical link" between collection of information and the explanation of its meaning. Code is a 'research-generated construct' and is an interpretation of the data for purposes such as pattern detection, categorization, theory building and so on (Saldaña 2015).

In this research, the principles of open coding were prevalent during the coding and analysis of collected data. Open coding has been defined by (Strauss and Corbin 1998) as "analytical

process through which concepts are identified and their properties and dimensions are discovered in data". After transcribing the interviews, they were analyzed in depth by going through the text and marking the common concepts coming up during the reading. These common repetitive concepts across the interviews were then categorized, and these categories were decided by the author based on the research topic. Although a large number of categories could have been drawn out from each transcript, it was decided to narrow down the number of themes due to paucity of time for the analysis.

From the analysis, four themes relating to the challenges faced in energy data management were discovered. They are – existing collection mechanism, data gaps, socio-political motives and lack of research in the field. These have been discussed in detail in the following chapter.

3.5 Limitations of the Research

There were multiple limitations and challenges that the research faced. Firstly, due to the time of the year (Summer), many interviewees were on vacation and travelling. This made it difficult to establish contact with them and request for their time. Some of the participants of the study were experts working in the governmental bodies like ministries. Due to strict schedule and confidentiality, the access to them was difficult. Hence, it took a longer time to establish contact with them and fortunately, most of them agreed to participate in the research. It is often the case that employees working at ministries don't respond over the emails. Hence, some of them were contacted through social media website like Twitter and LinkedIn.

4. Energy Data Management in India – Results and Discussion

4.1 Introduction

In this chapter, I will be explaining the research questions posed at the very onset of the study and review a few dominant themes which emerged out of the data collection process. While doing so, I will also attempt to seek relevant resemblances with the existing body of literature around the topic. Hence, the chapter contains results and discussions of the collected data to show its significance and relevance to the research questions.

The field work (interview) was performed between 15th May and 3rd June. The location of the interview was generally the office of the interviewee. In total, ten people were interviewed while more were approached for it and the average duration of an interview was between one hour to ninety minutes. During the course of this chapter, I will be using thematic analysis to review the data collected during the interviews.

4.1.1 Background of interviewees

In the beginning of the interviews, consent about the study was taken and the measures about the anonymity were informed. Even though most of the respondents agreed to usage of their names, I decided to use anonymity on all to be consistent during the study. Prior appointment was fixed with interviewees through face-to-face request, emails and phone calls. When none of these mediums of communication worked, social media was used to connect with the experts. For example, one of the experts leading a government body was contacted through Twitter, while a person leading private solar PV agency was contacted through LinkedIn.

The interviews were taken from experts working in the field of energy to better understand the nitty-gritties of the system and the challenges faced by them. The interview questions were drafted to be 'humanistic' (Mann 2017), exploratory and open-ended. All of the respondents

come from an educated background, or in other words, are economically privileged group with Indian origins. In these ten interviewees, three were female and seven were male. The nature of their employer differs, and it covers a wide range – government, university and private research institute. In a qualitative research, the individual description of the participants is more important than the overall description of the group. I have included brief background of some participants with relevant information in the following paragraphs.

In India, the major governmental entities responsible for maintaining energy data are NITI Aayog, MoSPI and CEA. Hence, it made sense to take interviews from at least one resource person from each of these bodies who are involved in the energy data management process. Due to strict schedule and security at these government offices, it is difficult to get in touch with the right person and one has to scrape a lot of information online and try to establish contact from one way or the other. The detailed interviews with them provided a lot of insights in how the system in government works and the future plans and up-to-date information about the developments. One person from a private organization dealing with solar PVs was also interviewed.

There was equal representation with three participants from academia. One of the representatives from private research university in India works with the management of energy data in transport sector. The primary type of data required in their research is activity data, meaning the data about movement of either passengers or freight. It is also essential to know what kind of energy data they deal with and that is how every interview started. When asked on how they procure such data and make use of it:

"we rely on registered vehicle data which is published annually by ministry of road transport. Then by factoring in the GDP and population growth rate, we determine the approximate stock on the road. We also carry out surveys to what kind of vehicle utilization is happening, age of the vehicle, what are the efficiency levels and fuel usage."

Three researchers from think-tanks specializing in energy research were also interviewed for the collection of primary data. These think-tanks have special group on energy policy research and typically work independently. One of them shared their primary source of data collection:

"Currently ministries are collecting data and publishing it for our consumption, which we further use to draw our insights."

4.1.2 Energy data management: Indian context

India's EDM system is relatively decentralized when compared to the United States or the UK (Liu, et al. 2017). There are multiple agencies which handle different processes leading to a scattered spectrum of energy data. Three type of institutions involved in the management – energy line ministries and their sub-ordinate agencies, statistical agencies at both central and state level, and finally, the planning institutions (Liu, et al. 2017). Therefore, there is no one central authority responsible for all the functions that the management of energy data demand.

There are scholars who have engaged insightfully with the energy research in India and knowing their perception about the energy data management in India made sense to better grasp the overall picture on how the perceptions are divided.

It is tempting to compare and draw models from the already developed systems and apply them to the other system. While similar cases have been happening in India from quite long, the results which the international agencies provide do not convince the researchers in the country. The participant from NITI Aayog's energy specialization group mentioned this reason when asked about views on international collaborations in the past:

"First of all…The case of India is very different here because we have five ministries for energy which makes the system scattered."

While trying to strike a balance between the pros and cons of the system, it was evident from the interviews that the entire sample group agreed on the quality of supply side data. The report

on 'Assessment of Energy Data Management in India' also reported that the energy supply data is '*reasonably good though there is a room for improvement*' (Dukkipati, Iyer and Sreenivas 2014). One participant from a private research institute specializing on energy aptly described the good side of the supply side data and the collection system:

"The good thing is that supply side data is extensive and presented in many forms. One of the best model systems for thermal power systems is CEA, however, all of them are in pdf form. Another good thing is that since there are different departments publishing data, we can easily cross check the data and makes our work credible."

The participant from CEA also mentioned about their experience with the supply side data:

"There are three type of generating entities in India - central, state and independent power producer. Irrespective of ownership, we (CEA) collect the data from them and publish daily reports. Up to 33 KV level, the data visibility is very good."

4.2 Challenges to the current energy data management system

During the course of data analysis, I came across four dominant themes relevant to answering the research questions. These have been discussed with original quotes from the participants of the study in the text following.

4.2.1 Existing data collection mechanism

Given the highly decentralized nature of the India's energy data management system, the critical key to the uniformity is the coordination between all the agencies. The problems in existing energy data management mechanism were discussed across all the interviews. Handling of data by the line ministries should be geared towards public dissemination but currently it is merely an administrative formality, notes (Liu, et al. 2017). For example, although daily data is published, it is usually in the form of pdf files which makes it difficult for the researcher to access the data.

The study by Liu and others (Liu, et al. 2017) suggest that the current data collection practices in India are consistent with the international practices. There are administrative provisions and legal mandates which empower the agencies to collect data through various mediums like census, surveys, automated sensors and so on. However, these data types are dominated by supply side data, which is a more formalized sector. The participant from a government body made similar conclusion about the availability of energy data:

"In our (NITI Aayog's) consultations, we have realized that the supply side data is available in many cases and its relatively well collected. The only problem there is the collation of who's going to do what? And there is data being collected but we don't know that it is being collected which we found out through our consultations. So, supply side data is not hard to get, it is there. It's just a matter of how its collected and on what periodicity and who published it."

It was clear from all the interviews and literature review that the current data collection mechanisms produce very scattered data. In one of its report on data management, the National Statistical Commission documented about the challenges to conduct surveys (National Statistical Commission 2011). The report mentioned that the resources provided to expand existing surveys or to introduce new surveys is very limited. In the only case study on India's energy data management, researchers Liu and others (Liu, et al. 2017) report that the fundamental body for the coordination – MoSPI doesn't have adequate authority and expertise in the role. A participant from government agency suggested that there are datasets which aren't published:

"MoSPI only has the mandate to publish such data, but the other ministries and departments might have useful data which we don't currently know about."

The rapid advancement in the technology change can be one of the reasons that might soon deem the current collection mechanisms to go outdated. One of the respondents from a private research institute had to say:

"The requirements of data have changed, and you have to also change the data collection process which meets the requirement."

Another concern that came across in the interview about the current system is the rapid switch from conventional sources to renewables. The current structure within the system is scattered while there is one government body CEA which handles electricity data irrespective of the source of generation. Participant from a private university in India voiced about the sources of data collection for their research:

"We (transport sector) have to deal with ministry of petroleum and natural gas a lot for the collection of data points at the moment but the government is rapidly pushing for electric mobility in the future."

Surveys are an important instrument used to collect data and solves the purpose of collecting extra information which is not published usually. The same participant mentioned a study by a private firm to gauge consumption patterns within the road modality by surveying costumers at the petrol stations. However, these techniques of studying consumption patterns might become even difficult when electric vehicles come into the picture. The electric vehicle owners have an option to charge their batteries at home or at office, which might go unaccounted if such surveying methods are not updated with time.

"In 2013-14, Petroleum Planning & Analysis Cell (PPAC) conducted a survey with Nielsen at petrol stations in the country. If our government departments can do these general surveys routinely, then there won't be any need for us paying extra for such additional studies."

The term 'Big data' has become very popular in the last years and is referred to the huge volume of data which has the potential to provide valuable information if proper analysis is operationalized (Lovell and Watson 2019). The invent of new technologies with modern sensors provide high quality datasets, which were once a luxury to obtain. Some of these new technologies are smart digital meters, substation sensors and even the household items

connected to the network with 'Internet of Things' (Lovell and Watson 2019). Widespread usage of such enabling technologies with networked sensors would provide immense amount of data. However, it is up to the agency as how they make value out of such data.

The creation of a central entity, similar to EIA of the US has been around in the system with scholars like Tongia and others (Tongia, Rai and Shrimali 2017) (Rai, Tongia, et al. 2017) (Chikkatur and Chakravarty 2008) been advocating for it. In the discussion about this topic, the interviewee from a government agency notes if such an agency is formed, it will be formed under NITI Aayog, which however, would be added as an independent department in the MoSPI :

"In the recent years, they have realized that the data is important and there are ideas floating around about the creation of an energy agency like US EIA. Talks of such sort are happening."

"From what I see, even in the long run MoSPI is going to be the ultimate agency but it just has to improve the way it publishes the data."

"Initially will be housed under NITI Aayog and will be associated body under NITI Aayog. Till such a time comes when it gets its own secretariat on functioning…because giving it to any energy ministry will be very biased given that all of them require it."

4.2.2 Data gaps

Data gaps has been defined as "*data useful for policy research, formulation, program monitoring, or decision making by various stakeholders, but not accessible in the public domain at desired granularity or frequency in a convenient form"* (Prayas Energy Group 2015). During the course of interviews, the same factors which make the data useful were mentioned repeatedly.

Liu (Liu, et al. 2017) mentions the issue with demand side data in India. A limited amount of data is available through the census and National Sample Survey Office by MoSPI. While surveys are an important tool to find information, the energy data collected through surveys in

India are spread over many questionnaires and not to inform the analysis in general. Furthermore, the surveys might be conducted at different time frames and this could lead to difficulty in processing coherent data altogether. The participant who specialises in the field of energy data management in the transport sector noted:

"We primarily look at the activity data related to movement. This is primary input we need to reach the final energy consumption calculation. In order to capture these parameters, there are a lot of data gaps within the transport sector which leads to inaccuracies in these basic inputs."

Similarly, (Kar, et al. 2019) conducted a study looking at the sales data of cooking gas cylinders to assess impact of a rural upliftment scheme in India. The research suggests a mid-course revision in the policy of the government. The datasets used in the study have a limitation that they don't collect demographic and socioeconomic data at the individual and household level, which limits the scope of the study and prevents the researcher to unfold another layer with the research. Although there was a response which looked at the scattered data from a different perspective. The availability of data at numerous agencies in different forms enables cross verification of the assumptions or projections, and even the validity of the research. When asked about how they realize data gaps, one participant from a private university mentioned:

"We cross check and perform energy balance to validate the results and realize data gaps."

Although the supply side data collection system is robust, emergence of new technology has made a loophole already visible. One expert from a government body provides information about the current gaps in supply side energy data:

"Except for the renewables, particularly off-grid, the supply side is there but just needing to be collated and analyzed."

The energy data about households is usually reported in the frame of how many of them have been connected to the grid, or in other words, are electrified. However, the recent trend of

households leaving the grid connection and switching to off grid technologies does not get monitored and has been overlooked (Lovell and Watson 2019). In their research 'Scarce data: off-grid households in Australia' aimed at exploring motivations to go off grid, they found out that factors like better finances, environment and social responsibility, and feeling of empowerment were among the most noted reasons. They concluded their research stating that the

"Lack of data in this instance results in a situation where off-grid issues have been left out of the energy policy frame: off-grid is not seen as relevant, because it is not visible as a policy issue" (Lovell and Watson 2019)

Scarcity of data might result in affecting the governance with drawing more attention to the policy areas where more data is available. (Huges 1983) in their book 'Networks of power: electrification in Western society' mentions that such a situation would favor the existing energy institutions and technologies while hindering the growth of innovation. The popularity of off-grid solar PV has been increasing over the past few years as the rates have come down and the technology has become more reliable. However, there is no mandate to collect the data from such off grid solar PVs which are smaller than 1 MW in generation capacity. A government expert shared what one DISCOM in Rajasthan has being doing to tackle this problem:

"An online portal has been created by one DISCOM in Rajasthan where all permissions are taken for the installation irrespective of capacity. They install their own meters to see how much energy is being generated. We are trying to study that."

Although is not important to capture the off-grid energy data at a 100% accuracy, but still it is essential part for the energy balance. Such negligence in the legal mandates could prevailing

because the off-grid data is not available, which makes the sector less visible. However, one participant mentioned about the increasing acceptance and growth rate of off grid PVs in India:

"There are a significant number of units less than 1 MW because it is very easy to install."

"A lot of these houses (with solar PV under 1 MW) do that (install PV) and are also connected to the grid. Their electricity consumption data fluctuates since they might pull in from the grid only once in a while. In a larger scheme, we don't know why. Right now, we just have to speculate."

4.2.3 Socio-political motives

The electricity sector in India has already witnessed a dual-track economy, meaning a state-run production exist side-by-side with market run production. This situation has resulted in the industrial costumers leaving the state-run system and installing their own power generation set up due to overly increased tariffs by the state government, as noted in a research named 'The politics of power: Electricity reform in India' by Kelli Joseph (Joseph 2009). Major fluctuations in the energy subsidy policies can spark public protests and eventually affect the regime stability, which is the reason why major oil and gas producers provide subsidies to their residents. An expert from government agency explained:

"Power is concurrent subject, so policy formation is in the hands of central government, but operation is within the purview of state government. In the data, same issues come because we cannot force them to share everything. We get the data which is required for policy planning and operation purpose."

Household electrification rate plays an important role in determining the economic as well as human development rates. India has been long implementing rural electrification projects to provide electricity at the remotest locations. Making the electricity accessible and provisioning affordable electricity can improve the legitimacy of the government in power. The data about electrification and the tariffs can lead to an additional variable in political circumstances, if left for open access (Shaffer 2017). Expert from a government agency explained how state governments manipulate the data for fulfilling their promises:

"Because each state's performance is being measured by a parameter called 'shortage' earlier, but we changed the name to 'power not supplied'. At the grid level sufficient power is available but due to their own commercial conditions or to reduce commercial losses due to theft, states are not procuring that much electricity. Their state government don't want to reflect it on their reports. Hence, there is a gap between load shedding in the power system. States have started unscheduled power cuts which they don't report. For example, some states promise 20 hours of supply to some area and when such is fulfilled, they report it as 'no shortage' but actually there is a shortage of 4 hours. This is the blind spot which we are not able to bridge."

On the similar line, ministries which might not benefit with the new decisions usually show reluctance in supporting the initiatives. A participant from the same government body shared their experience about a meeting in their group:

"Another problem is that the ministries might not work in tandem. All the things have to be coherent; you can't say different things. Each ministry wants to promote its own fuel and they have their own views. For EVs, we had a meeting, but the petroleum and natural gas ministry never shows any enthusiasm for it and rarely come to support. It is very subtle."

Studies about the politics of reform usually focus on the role of organized interest groups like agriculture in supporting or becoming an obstacle in the reform efforts (Joseph 2009). Some consumers in the country receive subsidized electricity and in order to recollect the costs incurred, the State Electricity Boards start to charge other users more than the actual price (Joseph 2009). These other users which get charged more than normal pricing are generally industries which fall in this trap. This happens when leaders of political parties support and vocalize the agriculture sector who don't want to purchase electricity at the market price Such politics renders the already existing regulatory agencies at both regional and state level as ineffective and of no use (Joseph 2009). An expert from a government agency shared their experience about the state agendas:

"When it comes to electricity, the states have their own political agenda. Sometimes they refrain from giving proper data thinking it might affect the state allocations. When politics come into this, we have something called free electricity… for farmers, this and that. This kind of defeats the whole purpose of data we have because it is not really helping policy making.

The moment you decide to give free electricity, then no use of us sitting here and analyzing data just for it to be rained upon in one moment by politician who comes under power. It changes the whole dynamics of the system."

The analysis of distributional coalition by Olson (Olson 1982) and Przeworski (Przeworski 1991) applies to the group which is most vulnerable to the reforms and when the politicians decide to raise tariffs, there have been classic examples of them voted out in the next election term (Joseph 2009). Therefore, politicians are left with little choice to support pricing reforms. (Shaffer 2017) notes the same about energy subsidies in their paper.

In addition to the problems created by the distribution at low and subsidized costs to a certain organized sector of the society, electricity theft is another reason which distorts the electricity sector in India (Joseph 2009). The low-cost electricity and high level of theft renders the electricity board economically weak which, therefore, limits the much-needed infrastructural improvements (Joseph 2009). As an interviewee from a government body note:

"The electrification rates are for those households which are eligible for it. What about the 'kaccha' (temporary) houses? Technically they are also consuming electricity but are not on meter. So, do you consider it as consumption or loss? There is no way to differentiate that but it still is a big part. The minute you consider it as a transmission and distribution loss, you think about investing in the infrastructure for transmission. It is technically not going to work because that loss is actually consumption. Electricity theft is still a big problem."

4.2.4 Lack of research in the field

Studies on energy data are predominantly about the economics of switching to renewables, performance analyses, modeling of future scenarios and so on. These studies are most commonly built on a combination of technical as well as other quantitative parameters. For example, a study by (Khalilpour and Vassallo 2015) 'Leaving the grid: An ambition or a real choice?' explicitly suggests the requirement of a socio-behavioral study for the said topic which

investigates ‘only’ the economic feasibility of switching off the grid. One of the researchers who participated in the interview shared their experience about this field:

“I am glad I got to put out my points. These are the things we still have to look into.”

Due to limited research in Indian context, the results of international frameworks create a mismatch when applied to India. The results published by international agencies often lack certain parameters and are based on assumptions. One of the researchers wasn’t convinced with a report published by IEA:

“What we have seen is that there is a lot of mismatch between the international and national reports. I had a discussion with IEA for my past project and I wasn’t convinced with their results because they had a lot of deviation from what we have.”

Even the experts at government bodies mentioned that there is very limited interest when it comes to energy data management, in contrast to energy tariffs. One reason for the limited research could be attributed to lack of public participation in this field of interest. The participant at the government agency shared their experience about the lack of public as well as academia participation:

“When we put out public meetings for suggestions, there is almost no participation from general public.”

“Feedback from researcher is not coming to us. I see a lot of such research happening in Germany, but none in India. Every time I have to make some decision, I have to consult with private organizations, and they produce reports using international data and give me a policy recommendation. Academia and industry should provide comments and feedback to us. They are out of touch with the industry and the reason for this is utilities are not sponsoring research. For example, new projects in the US are supported by industry and it is a win-win situation with a collective goal.”

4.3 Limitations

There could be some limitations on the generalizability of findings of this study. However, the attempt was made to include all the relevant stakeholders and also to create a mix-gendered sample group for interviews. The results would have been more holistic if the study could have included experts from the industrial sector. In addition to them, experts leading new business models and innovations like start-ups which are purely utilizing energy data for their revenue or service could have served fresh narratives to the study. However, finding such entrepreneurs was very difficult. One reason for the limited number of such ventures could be the ineffective data management scene in India which makes it less reliable for the people to depend on it for revenue purposes.

5. Conclusion and recommendations

Energy data plays the role of a critical enabler for people involved in policy formulation and research. Although there are weaknesses in the energy data management system as discusses in the chapters above, it is consistent with the political and administrative structure by being decentralized in nature.

In order to develop standardized norms and classifications, it is important the governing bodies establish some common principles which are shared as a core between all the responsible bodies involved in the energy data management. This would, in turn, enhance coordination between the agencies. NITI Aayog is leading the currently involved agencies to come together under one umbrella for meaningful discourse to maximize the value of ongoing efforts. Given the next census is coming, it is an opportunity to make the first holistic energy survey a part of it.

Since there is a vast amount of data that can be collected from different resources, it is worth to prioritize the data which is more useful and can fill in important data gaps, as mentioned in the study. Focus needs to be on the energy data crucial for planning and development. The data relevance can be improved by working closely with the relevant agencies, policymakers and users of the energy data. As one of the participants from a government agency right note:

"All kinds of energy data might not be useful from the perspective of planning, so we need to prioritize on what kind of data should we focus on, given on how many gaps are there in the system."

Innovation can play a key role in driving the data management forward. Automated technology with networked sensors can make the collection process of data quicker, easier and in machine readable format for the publication without any manual errors. For example, there have been

cases such as electricity theft or under-reporting of the usage. Use of such technology would also ease the process and quality of data dissemination. The new data dissemination portals can also make use of such new technologies. Rather than making these initiatives a one-time exercise, efforts have to be made so that it happens in continuity. One suggestion from the expert at government body mentioned that there is no feedback mechanism for the data dissemination websites. In order to make these websites useful and not a one-time affair, the process to provide feedback has to be streamlined.

If the end goal is to go electric, then all the ministries would require proficiency in electricity data collection in their departments and this would therefore increase the demand of professionals required in the field as well. Ideas about creation of a new central entity solely responsible for energy data management has been floating in the system. Researchers like (Tongia, Rai and Shrimali 2017) (Rai, Tongia, et al. 2017) (Prayas Energy Group 2015) (Dukkipati, Iyer and Sreenivas 2014) have mentioned the importance and advocated for such a central entity in their work in the past.

The opportunities that off grid solar brings to the plate are enormous but datasets are required to be able to make use of its full potential. Currently, CEA collects unit wise data from every generation station except renewable, because they mainly deal with conventional sources and renewable over 25 MW generation capacity. Hence, there is a division between the responsibility between the CEA and MoSPI which further makes the case complex.

The research attempts to introduce the view of critical data studies in the energy studies, in the hope to advocate greater use of qualitative social science in the energy data research, similar to the developments in the big data research. Such an approach is valuable because it pulls our attention to the ‘power’ and ‘influence’ of missing or non-existent data which is not neutral in its effects (Bowker and Star 2000).

6. Bibliography

Tongia, Rahul, Varun Rai, and Gireesh Shrimali. 2017. "India needs agency for energy data." *Correspondence*, January 5: 30.

Rai, Varun, Rahul Tongia, Gireesh Shrimali, and Nikit Abhyankar. 2017. "Data for development: The case for an Indian energy information administration." *Energy Research & Social Science* 105-109.

Ministry of Law and Justice. 2003. *The Electricity Act, 2003.* New Delhi: Ministry of Law and Justice, June 3.

Lovell, Heather, and Phillipa Watson. 2019. "Scarce data: off-grid households in Australia." *Energy Policy* 502-510.

Laws, Nicholas D., Brenden P. Epps, Steven O. Peterson, Mark S. Laser, and G. Kamau Wanjiru. 2017. "On the utility death spiral and the impact of utility rate structures on the adoption of residential solar photovoltaics and energy storage ." *Applied Energy* 627-641.

Singh, Kartikeya. 2016. "Business innovation and diffusion of off-grid solar technologies in India ." *Energy for Sustainable Development* 1-13.

Spangher, Lucas. 2019. "A case study comparison of different ownership and finance models for off grid systems in southern india ." *Energy Procedia* 472-477.

Huges, Thomas P. 1983. *Networks of Power: Electrification in Western Society 1880-1930.* Maryland: The John Hopkins University Press.

Christensen, Jesper Lindgaard, and Daniel S. Hain. 2017. "Knowing where to go: The knowledge foundation for investments in renewable energy." *Energy Research & Social Science* 25: 124-133.

Iyer, Maithili, Satish Kumar, Sangeeta Mathew, Hannah Stratton, Paul Mathew, and Mohini Singh. 2016. *Establishing a Commercial Buildings Energy Data Framework for India: A Comprehensive Look at Data Collection Approaches, Use Cases and Institutions.* Research Study, University of California, Berkeley: Lawrence Berkeley National Laboratory.

EPA. 2018. *Centralized Generation Of Electricity and Its Impacts on the Environment.* March 13. Accessed June 13, 2019. https://www.epa.gov/energy/centralized-generation-electricity-and-its-impacts-environment.

Pepermans, G., J. Driesen, D. Haeseldonckx, R. Belmans, and W. D'haeseleer. 2005. "Distributed generation: definition, benefits and issues." *Energy Policy* 33 (6): 787-798.

Ackermann, Thomas, Göran Andersson, and Södera Lennart. 2001. "Distributed generation: a definition." *Electric Power Systems Research* 57 (3): 195-204.

Sovacool, Benjamin K., and Torben Tambo. 2016. "Comparing consumer perceptions of energy security, policy, and low-carbon technology: Insights from Denmark." *Energy Research & Social Science* 11: 79-91.

Dukkipati, Srihari, Rakesh K. Iyer, and Ashok Sreenivas. 2014. *An Assessment of Energy Data Management in India.* Research Study, Pune: Prayas (Energy Group).

Ministry of Power. 2019. *About Ministry.* June 20. Accessed June 20, 2019. https://powermin.nic.in/en/content/about-ministry.

Pal, Animesh. 2013. "POWER SECTOR IN INDIA: GROWTH, POLICIES AND CHALLENGES." *International Journal of Emerging Technology and Advanced Engineering* 3 (3): 527-536.

Dalton, Craig, and Jim Thatcher. 2014. *Society + Space.* May 12. Accessed June 20, 2019. http://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher/#.

Kitchin, Rob, and Tracey Lauriault. 2014. "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work." *The Programmable City Working Paper 2.* Ireland.

Prayas Energy Group. 2015. *Data Gaps in India's Energy Sector.* Indo-US energy dialogue, Pune: Prayas (Energy Group).

Chikkatur, Ananth, and Shoibal Chakravarty. 2008. "Need for an Integrated Energy Modelling Institution in India." *Economic and Political Weekly*, May 24: 64-69.

NITI Aayog. 2019. *Data Management.* May 10. Accessed May 10, 2019. https://niti.gov.in/content/data-management.

Sankar, T L. 1985. "Energy Policy Formulation - The Indian Experience." In *Public Policy and Policy Analysis in India*, by R. S. Ganapathy, S. R. Ganesh, R. M. Maru, S. Paul and R. M. Rao. New Delhi: Sage Publications.

Thollander, Patrik, Svetlana Paramonova, Erwin Cornelis, Osamu Kimura, Andrea Trianni, Magnus Karlsson, Enrico Cagno, Inés Morales, Juan Pablo, and Jiménez Navarro. 2015. "International study on energy end-use data among industrial SMEs (small and medium-sized enterprises) and energy end-use efficiency improvement opportunities." *Journal of Cleaner Production* 104: 282-296.

CEA. 2013. *Report of 18th Electric Power Survey of India.* Power Survey Report, New Delhi: CEA.

Shaffer, Brenda. 2017. "The Potential of Energy Data ." *Central Intelligence Agency.* December 1. Accessed July 2019. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol-61-no-4/pdfs/energy-as-analytic-tool.pdf.

Mukherjee, Shilpi, Tarun Dhingra, and Anirban Sengupta. 2017. "Status of Electricity Act, 2003: A systematic review of literature." *Energy Policy* 102: 237-248.

IEA. 2019. *About us.* July 20. Accessed July 20, 2019. https://www.iea.org/about/.

Iyer, Maithili, Satish Kumar, and Sangeeta Mathew. 2016. *Establishing a Commercial Buildings Energy Data Framework for India: A Comprehensive Look at Data Collection Approaches, Use Cases and Institutions.* Recent Work, Berkeley: Lawrence Berkeley National Laboratory.

Liu, Bo, Meredydd Evans, Sha Yu, Volha Roshchanka, Srihari Dukkipati, and Ashok Sreenivas. 2017. "Effective energy data management for low-carbon growth planning: An analytical framework for assessment." *Energy Policy* 32-42.

OECD. 2019. *OECD Data.* July 31. Accessed July 31, 2019. https://data.oecd.org/energy/electricity-generation.htm.

Phadke, Amol. 2014. *Avoiding 100 New Power Plants by Increasing Efficiency of Room Air Conditioners in India: Opportunities and Challenges* . Environmental Energy Technologies Division, Berkeley: Lawrence Berkeley National Laboratory .

Rai, Varun, Rahul Tongia, Gireesh Shrimali, and Nikit Abhyankar. 2017. "Data for development: The case for an Indian energy information administration." *Energy Research & Social Science* 105-109.

Tongia, Rahul, Varun Rai, and Gireesh Shrimali. 2017. "India needs agency for energy data." *Correspondence: Nature*, January 5: 30.

Saldaña, Johnny. 2015. *The Coding Manual for Qualitative Researchers. 3rd edition.* Los Angeles; London: SAGE Publications Ltd.

Charmaz, K. 2001. "Grounded theory." In *Contemporary field research: Perspectives and formulations (2nd ed.)*, by R. M. Emerson, 335-352. Prospect Heights, IL: Waveland Press.

Strauss, Anselm, and Juliet Corbin. 1998. *Basics of Qualitative Research: Second Edition: Techniques and Procedures for Developing Grounded Theory. 2nd Edition.* Thousand Oaks: SAGE Puvblications, Inc.

Mann, S. 2017. "Big Data is a big lie without little data: Humanistic intelligence as a human right." *Big Data & Society* 1-10.

Kar, Abhishek, Shonali Pachauri, Rob Bailis, and Hisham Zerriffi. 2019. "Using sales data to assess cooking gas adoption and the impact of India's Ujjwala programme in rural Karnataka." *Nature Energy* 1-9.

Joseph, Kelli L. 2009. "The politics of power: Electricity reform in India." *Energy Policy* 38: 503-511.

Olson, M. 1982. *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities.* New Haven: Yale University Press.

Przeworski, A. 1991. *Democracy and the Market.* Cambridge : Cambridge University Press.

Khalilpour, Rajab, and Anthony Vassallo. 2015. "Leaving the grid: An ambition or a real choice?" *Energy Policy* (Energy Policy) 207-221.

Bowker, G.C., and S.L Star. 2000. *Sorting Things Out: Classification and its Consequences.* Cambridge, Massachusetts: The MIT Press.

National Statistical Commission. 2011. *Report of the Committee on Data Management.* Report, New Delhi: National Statistical Commission.