DEVELOPING A FRAMEWORK TO ASSESS SOCIO-ECONOMIC VALUE OF OPEN DATA IN INDIA

Sharon Buteau
IFMR LEAD
Chennai, INDIA
+91-9600058993
sharon.buteau@ifmr.ac.in

Preethi Rao
IFMR LEAD
Chennai, INDIA
+91-9884717056
preethi@ifmr.ac.in

Anshuman Kr Mehta
IFMR LEAD
Chennai, INDIA
+91-9791478547
anshuman.mehta@ifmr.ac.in

Vigneshraja Kadirvell
IFMR LEAD
Chennai, INDIA
+91-9597639410
vigneshraja.kk@ifmr.ac.in

ABSTRACT

This paper attempts to develop a framework to assess the socio-economic value of Open Data in India. The paper discusses the various actors, their roles vis-a-vis usage of Open Data to generate intended output, and the net results of expected outcomes at the macro level. The paper further elaborates on how the framework is intended to observe and measure benefits arising as a result of Open Data production and utilization across various sectors in India and the value it creates for the stakeholders. The framework developed in this paper is intended to form the basis of a more elaborate study under which we aim to determine the rupee value of Open Data in India. An assessment of the value created by Open Data will provide the necessary insights to Open Data producers to base their decisions as regards scaling of their efforts as well as provide the much necessary feedback to the Open Data ecosystem as a whole.

Keywords

Open Data, Socio-economic impact, OGD – Open Government Data, Cost-benefit analysis

ACM Reference format:

S. Buteau, P. Rao, A. Mehta, V. Kadirvell 2018. DEVELOPING A FRAMEWORK TO ASSESS SOCIO-ECONOMIC VALUE OF OPEN DATA IN INDIA. In Proceedings of The 14th International Symposium on Open Collaboration (OpenSym '18). ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3233391.3233532

OpenSym2018, 08/22 2018, Paris

ACM ISBN 978-1-4503-0000-0/18/06
https://doi.org/10.1145/1234567890

OpenSym '18, August 22--24 2018, Paris

978-1-4503-5936-8/18/08...$15.00
https://doi.org/10.1145/3233391.3233532

1. INTRODUCTION

Open data is data that is not subject to privacy, security or privilege limitations. Data or content is open if anyone is free to use, re-use or redistribute it, subject at most to measures that preserve provenance and openness (The World Bank, 2017). As Open Data gains the status of a resource more valuable than oil in current times (The Economist, 2017), governments and businesses across the globe are gearing up to devise means to exploit its full potential. Over the past two decades, numerous studies sanctioned by international governments as well as cross-country unions have attempted to assess the economic, social and political value of Open Data and how best it may be capitalized. Seventy countries have established a multilateral network for strong Open Data commitments designed to foster greater transparency, generate economic growth, empower citizens, fight corruption, and more generally enhance governance, as a part of their National Action plans. Nearly half of these countries are developing countries (Young, 2017). This is evident of the fact that the concept of Open Data is not confined to developed countries alone but is also garnering attention in developing countries.

Over the last decade, a lot of work has been done on Open Data pertaining to economically advanced countries, such as estimation of net benefit occurring to the particular country as a result of Open Data such as through improved governance, through greater transparency, through incubation of new ideas in the form of data driven start-ups etc. and the empirical estimation of the impact of Open Data on wages, employment, growth rate and GDP. However, there are gaps in the existing literature on the socio-economic impact of Open Data in developing countries, especially India. Policymakers and data producers in general often ask the question such as; what is the scope of Open Data in a country like India? Open Data brings economic growth and social equality through employment generation, increased salary and so on. Is it really possible in India? If yes, how? This paper attempts to answer the above questions using a framework that has been adapted from studies undertaken in similar contexts.

2. BACKGROUND

This section deals with the existing literature, approaches and model on the measurement of socio-economic value of Open

Data. In addition to that, this section also highlights the gaps in the research as well as the link between past research and current scenario along with the applicability of different frameworks in the context of developing countries.

2.1 EXISTING MODEL AND LITERATURE ON OPEN DATA

Over the past two decades, a number of attempts have been made to develop a comprehensive system of valuation of Open Data and its impact on various economies across the globe. Each paper or article, although influenced by existing research as well as guidelines set by authorities, has innovated and added to prevalent techniques to devise a methodology best suited to the region, industry or scope of the individual study.

For instance, a **top-down value-added approach** has been applied to evaluate the size of the open government data sector of the economy in the European Union and the United States of America (PIRA, 2000). Using case studies, the method estimates the economic value added to calculate both the supply and demand sides of open government data. In general, a 'top-down' approach considers the value of open government data through the resources devoted to generating it or using it (Bureau of Communications Research Australia, 2016). Estimating the **supply side** involves calculating the incremental investment cost to the Government for collecting and sharing the data. Estimating the **demand side** involves calculating any expenditure undertaken on the Open Data by users and re-users, reflecting the value they place on this data (Bureau of Communications Research Australia, 2016). The findings for Europe under this study were used to extrapolate the size of the information sector in USA to be around EUR 750 Billion, although some activities were unrelated to Public Sector Information (PSI).

A top-down approach is useful in cases where the aim of the research is to identify/estimate the future importance of open government data , as this approach estimates the total values of the current sectors that use open government data as an input, and provides an indication of the gross value of the open government data to the economy (DotEcon, 2006) (Bureau of Communications Research Australia, 2016) (Deloitte, 2013). On the other hand, a top down approach can suffer from drawbacks such as double counting of benefits which leads to overstating the value of PSI, particularly where the extent of alternatives to the use of PSI are not identified directly (DotEcon, 2006). Another major criticism for the approach is its tendency to over-attribute causality and generate biased estimates (Deloitte, 2013).

The costs and benefits associated with PSI have been estimated using **'The welfare approach'** and **'The returns approach'** (Houghton, 2011). The benefit to cost ratio is estimated as 5.3. The welfare approach might be interpreted as indicative of the lower bound impact. This approach uses consumer surplus to arrive at the estimation from data on revenue and the elasticity of demand. The returns approach includes the multiplier effects that are not captured in the first approach. The latter approach is the extension of modified Solow-Swan model (Houghton and Sheehan, 2009) to estimate the potential impacts of increased open access to research publications and data on social returns to R&D. The paper introduces accessibility into the standard model as a negative or friction variable and assesses the impact on returns to R&D through reducing the friction by increasing accessibility.

Using case study analysis (ACIL Tasman, 2008) explore the level of impact and extent of adoption of spatial information in the private and public sectors. The paper discusses two scenarios. Scenario 1 reflects the impact that can be confidently and verifiably quantified through the use of reliable statistics, existing literature and expert opinion. Whereas, scenario 2 is the estimated scenario which comes closer to reality (as distinct from that which we can confidently quantify). The direct impact of spatial information has been estimated under two scenarios using 22 sectors of the Australian economy. These direct impacts have been applied to a '**Computable General Equilibrium Model'** (CGE) to calculate the aggregate impact of spatial information on the Australian economy. The spatial information has contributed an estimated 0.6% to 1.2% GDP under the two scenarios, increased household consumption by between $3.57 billion and $6.87 billion on a cumulative basis, increased investment by between $1.73 billion and $3.69 billion on a cumulative basis and increased real wages by 0.60% and 1.12% respectively.

The CGE model addresses drawbacks in the simple benefit / multiplier methodologies. The model also allows analysis of changes in macro-economic aggregates which result from task-related changes e.g.: Gross Domestic Product, income, investment, wages and employment, as well as enable a view of potential changes to government revenues from taxation and other sources.

(DotEcon, 2006) uses a **'Bottom up Approach'** to show the net value of PSI in the United Kingdom. This method does not suffer problems of double-counting and over-estimation that are associated with the top-down approach. The paper focuses primarily, but not exclusively, on central government PSI holders, particularly 400 public bodies with an annual expenditure of £500,000 or more. It broadly examines the effects of converting raw information collected by PSI Holders into value added products and services, access to and pricing of raw information effects on competition between PSIHs and businesses selling value-added products, PSIHs' engagement in value-added activities and the effectiveness of the existing regulatory and legal framework. The **'Bottom up Approach'** considers the net value of Open Data to society today, associated with the current PSI products and services. This is the net willingness to pay for PSI (the maximum price someone would pay for PSI minus the cost of supplying it). The results indicate that the net value of PSI in the UK was around GBP £590 million per annum in 2005, with a potential value of around £1.1 billion per year (around 0.1% of GDP)

In contrast to the top-down value added approach which tells us the overall significance of open government data by detailing the size of the industry, the bottom-up approach, which uses productivity analysis, details the potential increases in the size of

the economy as a result of open government data usage or initiatives. A bottom-up approach can be applied to address questions around the future productivity benefits that could be derived from open government data (Bureau of Communications Research Australia, 2016)

According to a PwC report on data driven innovations, innovation has added an estimated $67 billion in new value to the Australian economy, or 4.4% of GDP, broadly equivalent to the retail sector's contribution (PWC, 2014). The report applies PwC's **Geospatial Economic Model**, which allows economic factors to be assessed at a granular level. The model divides Australia into 2,214 locations, each with a population of around 10,000 people. For each location, it then combines multiple data layers from 2001 to 2013 with forecasts for 2020. For each location the data driven innovation analysis applied is as follows:

- The economic output, calculated in a way that is consistent and reconcilable with the income approach of measuring GDP and Gross State Product (GSP) by the Australian Bureau of Statistics (ABS).
- The number of employees and businesses in each industry and by business size, overlaid on economic output to identify labour productivity.
- The level of innovation activity, derived by overlaying results of the ABS survey of innovation on the number of businesses based on its industry and size.

The MEPSIR (Makx Dekkers, 2006) report undertook a baseline measurement of PSI market in the European Union under the main data domains of business, geographic, legal, meteorological, social and transport information. These domains were valued based on conditions of availability, accessibility, transparency, accountability, non-discrimination and economic results using sub-indexes for each condition. Two methods were used to arrive at the value of the PSI market. Under the first method, data holders, data users and re-users were asked to provide estimates regarding the size of PSI market pegging its value at approximately EUR 26.1 Billion in 2006. The alternate method was based on total turnover i.e. as the sum of the turnover of all re-users minus the amount spent on acquiring the PSI. Under this method, the average value of the PSI market of Europe was estimated to be worth around EUR 27.6 Billion (Makx Dekkers, 2006).

Capgemini has estimated the value of OGD in the EU28+ to increase by 36.9% between 2016 and 2020, to a value of EUR 75.7 Billion by 2020. Over 25,000 Open Data jobs are expected to be created in this four-year period (Capgemini, 2015) (OECD, 2006). This estimation was based on four internally developed matrices: Market size and value added as percentage of GDP, Number of jobs created, cost savings for public sector and efficiency or productivity gains. While the first 3 matrices involve quantitative analysis, the final matrix is based on qualitative assessment of the social impact of OGD. Interestingly, the study also forecasts that by 2020, public sector cost savings will be approximately EUR 1.7 Billion due to the impact of OGD initiatives.

3. GAPS IN OPEN DATA RESEARCH IN THE CONTEXT OF INDIA

While a number of studies have been conducted on the socio-economic impact of Open Data, these studies are mostly confined to developed countries. Thus there is an enormous potential to exploit the opportunity in Open Data in the context of a developing country such as India. Following are the gaps in Open Data research in the context of developing countries such as India:

- Most of the studies in India are based on awareness and accessibility of Open Data. These studies are mainly cross-sectional studies. As a result, they do not capture the changes in various fields and factors over a period of time. This presents tremendous scope for longitudinal research in this field to comprehend changes in the long run and provide stronger insights on long- term impact.
- Many studies from developed country and developing countries used 'case study method' to show the net benefit from Open Data usage. However, a major drawback of these studies is that the result doesn't capture wider socio-economic impacts and at the same time the findings cannot be generalized to other fields.
- In a country such as India, where the problem of 'digital divide' still exists, one can use comparative analysis of socio –economic impact of Open Data in urban, semi-urban/peri urban and rural areas to understand how to address the problem and overcome inequalities.

4. OPEN DATA THROUGH THE LENS OF DEMAND DRIVEN APPROACH IN INDIA

The concept of Open data is still in a nascent stage in India. In a country like India where there is a scarcity of resources and proper infrastructure, a large volume of existing government data is still inaccessible in digital formats. The granularity of open government data in India has always been a concern, given its inability to meet the stakeholders' needs as it uses only micro-level data. There is also serious demand for geospatial data for visualizing and communicating issues as they exist on the ground. The unavailability of official map data due to the conservative map policy of the government and lack of interoperability in sharing this data has discouraged better planning, tracking progress in the 'real space' and pushing government to take remedial steps (Parihar, 2015). The cost incurred in publishing open data poses another impediment and is a crucial factor. This incongruity in demand and supply of open data advocates the need for a demand driven model which can act as a tailor made instrument suited to the needs of developing countries. Doing so would ensure that the allocation of resources in publishing those data is optimized and caters to the quadrants where the actual demand is, eventually fostering sustainability in the longer run. The demand driven model

entails three categories enumerated below (The Cambridge Group, 2018):

- **Current demand** reflects expressed needs and wants. Challenging the status quo and improvising per the needs of the consumers is the mantra to remain relevant
- **Emerging demand** reflects needs that are embryonic but being articulated – those that are held by a small but growing consumer segment
- **Latent demand** sits outside the box. It comprises of unrealized needs that are yet to be identified by the consumers but are potentially disruptive and highly transformational

Addressing each type of data demand would bring greater potential of minimizing the cost from the supplier side. Determining the current demand is not that colossal a challenge when compared to the other two. For example, one of the unique features of the Open Data portal in India is that citizens or users can demand a specific dataset from the government and others looking for similar data can endorse these requests. It then becomes mandatory for a department to release that data if 100 such endorsements are raised for a particular dataset (Parihar, 2015).

Taking cognizance of the current data demand is not sufficient. Tracing the latent and emerging demands, recognizing the trends early and envisaging the trajectory of the demand will play a key role in maximizing the socio economic benefits. The goal can be met only if it outweighs the cost incurred in publishing the data. Understanding the future of data demand from a data provider perspective in its formative state will turn the tables for developing countries. Through a deeper understanding of demand for a particular data, data providers such as government agencies, private firms, NGOs and other non-profit organizations can join hands with intermediaries such as data storage, services providers, app developing firms etc. to share information and fulfil ultimate users' demand. Most of the open data portals ensure tracking the frequency of download for a particular data set along with primary user identity. For a better cost benefit analysis, the broader approach should be to identify not only the primary user of the data but also underpinning the reason for using the same. Data suppliers should deploy precise tools to categorize users of the open data into groups based on the demand they want to satisfy; benefit they represent along with the motive.

5. FRAMEWORK FOR DETERMINING VALUE OF OPEN DATA

A cost-benefit analysis of making data 'open' and freely available to the public can be done to ascertain its rupee value. For this, the direct as well as indirect impact of Open Data on India's data domains i.e. sectors/activities/levels at which data is generated, maintained or used need to be quantified.

Data sets can be categorized into different data domains based on data generation, use or potential. This paper proposes the following domain classification for the Indian context, based on availability and use of data:

- Economics and Business
- Crime, justice and legislation
- Agriculture
- Education
- Healthcare
- Transport and tourism
- Meteorological
- Geographic

The figure below illustrates the framework of Open Data in developing countries. This framework describes the demand and supply side of Open Data, the mechanism through which the Open Data are transformed to usable output by individuals, businesses and governments and the application of these outputs that generates downstream and upstream signals which can finally lead to wider socio-economic impact.

The supply side of Open Data constitutes two parts; first, data supplied by government bodies and the second, data supplied by other than government bodies such as data from the private sector, academic institutions and so on.

The demand side of Open Data comprises of intermediaries and ultimate users. The intermediaries act as a mediator between the final consumers and the Open Data supplier. The intermediaries include data storage firms, app and software developers, and companies which provide advisory services to other businesses and firms. These intermediaries perform specific operations on raw Open Data such as data analysis, enterprise/ business mash-ups (blend of diverse data intended to provide specific output), value chain identification, research, app creation and other machine learning algorithms to create intended output on demand by the ultimate user. The ultimate user includes public sector, business, individuals and community. The ultimate users can also directly generate specific output according to their prerequisite by performing the same operations without going to intermediaries.

The output section in the above framework shows the expected results pertaining to the application of different mechanisms by the ultimate users and the intermediaries. The embracing of these outputs by the ultimate users can trigger downstream and upstream signals. These downstream and upstream signals are the changes which are visible at the ground level and at the macro level. The downstream signals are the value created by the practise of Open Data along the supply chain. For example, app development for farmers for specific crop cultivation can certainly increase their decision making process. Improving decision making ability can lead to an increase in the income of the farmers and improve other social mobility parameters. This is an upstream signal. Some of the downstream signals are: less red tape for business, accountability and transparency, awareness and efficient utilization of the resources. Similarly, some of the upstream signals are data driven start-ups, increased tax from business activity and improved policy making.

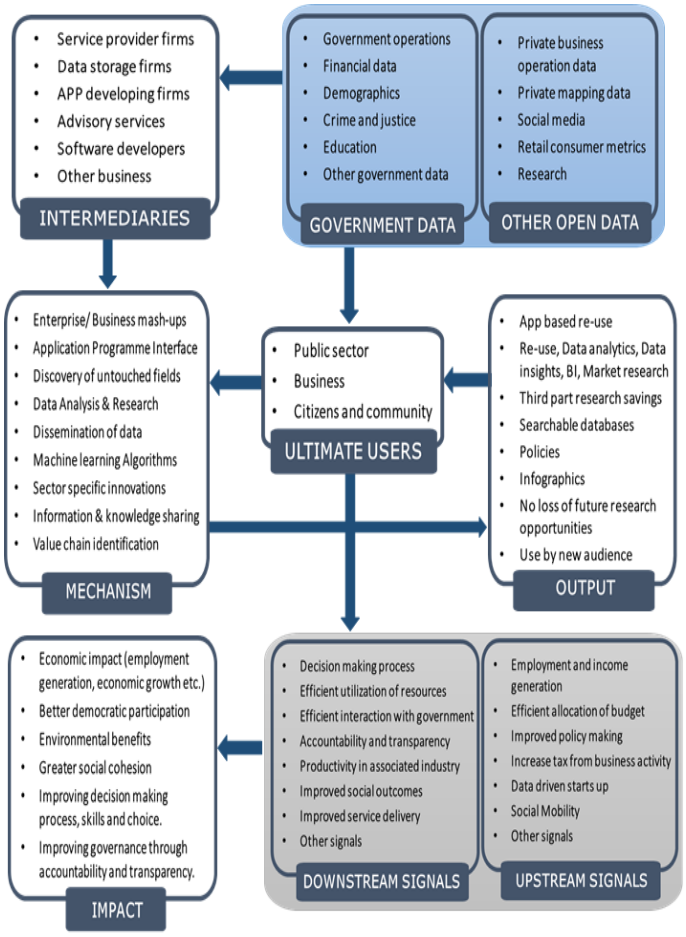

Figure 1: Framework for socio-economic benefit of Open Data in India

Finally, the cumulative effect of downstream and upstream signals can help estimate the comprehensive socio-economic impact as shown in the framework. For example, data driven start-ups may lead to wider economic impact such as employment generation and economic growth. Opening of crime, justice and legislation data will bring accountability and transparency which will further lead to better democratic participation and greater social cohesion. Similarly, downstream signals such as awareness about the application of green energy from apps or Open Data can lead to wider environmental benefits through adoption of sustainable environmental practices.

6. CONCLUSION AND FURTHER RESEARCH

Assessing the value of Open Data requires inputs and continuous feedback from all relevant stakeholders. The impact of Open Data is complex and multi layered, often trickling down to domains/activities that are not directly involved in storing or using data. In fact, the ultimate end beneficiary of Open Data is expected to be the entire society. The framework elaborated in this paper will form the basis of a larger study that we are currently working on in collaboration with the open government data team in India. The study intends to evaluate the socio-economic value created by Open Data in India.
Below are the broad objectives of the study:

- Identify the major domains in which Open Data is generated, stored and used and evaluate its impact on the specific parameters that represent them.
- Conduct quantitative and qualitative assessments of the tangible as well as intangible benefits resulting from the opening up of government data.
- Forecast impact of Open Data exploitation based on economic as well as social evidence.
- Develop, document and test a repeatable methodology to study impact of Open Data for specific sectors, industries or states etc.

For a developing country like India, where the influence of Open Data is not very evident, this study is one of the first few to attempt to estimate the value of its impact. Therefore, this study will require the counsel of specialists from a plethora of fields affected by Open Data, directly as well as indirectly. To ensure a comprehensive analysis, we have formed a consortium of experts to guide the research in the right direction and provide necessary insights. The members of the consortium represent a variety of domains, especially the data domains prioritized under the study. The consortium also has representation from state, central as well as departmental policy makers along with Open Data advocates and the private sector. Open Data is still in the nascent stage in India and future plans for developing and implementing Open Data projects will depend on the historic value creation and perceived benefits of opening up data.

References

[1] ACIL Tasman. (2008). The Value of Spatial Information: The impact of modern spatial information technologies on the Australian economy.
[2] Australia, B. o. (2016). *Open government data and why it matters: A critical review of studies on the economic impact of open government data.*Australia: Bureau of Communications Research—Department of Communications and the Arts.
[3] Capgemini. (2015). Creating Value Through Open Data – Study on the Impact of reuse of public data resources.European Data Portal.
[4] Deloitte. (2013). *Market assessment of public sector information.*Deloitte (UK) for Department for Business Innovation and Skills.
[5] DotEcon. (2006). *The commercial use of public information(CUPI).*London: Office of Fair Trading.
[6] Houghton, J. (2011). *Costs and Benefits of Data Provision.*Centre for Strategic Economic Studies Victoria University.
[7] John Houghton and Peter Sheehan. (2009, March). Estimating the Potential Impacts of Open Access to Research Findings. *Centre for Strategic Economic Studies Victoria University.*
[8] Makx Dekkers, F. P. (2006). Measuring European Public Sector Information Resources (MEPSIR).
[9] OECD. (2006). *Digital broadband content: Public Sector Information and Content.*Paris: Directorate for Science, Technology and Industry.
[10] Parihar, I. (2015). *How is open data changing India?*World Economic Forum.
[11] PIRA. (2000). *Commercial exploitation of Europe's public sector information.*Pira International Ltd, University of East Anglia and Knowledge View Ltd and European Commission, Directorate General for the Information Society.
[12] PWC. (2014). *Deciding with data How data driven innovation is fuelling Australia's economic growth.*PricewaterhouseCoopers for Google Australia.

[13] The Cambridge Group. (2018). Uncover Latent and emerging demand know where demand is heading before it is reflected in the market. *The Cambridge Group*.
[14] The Economist. (2017, July). The world's most valuable resource is no longer oil, but data.
[15] The World Bank. (2017). Open Data Essentials. *Data*.
[16] Young, S. G. (2017, November 13). Open Data in Developing Economies: Toward Building an Evidence Base on What Works and How. *African Minds*.