Open Data and Applied Socio-economic Research in India: An Overview

Sharon Buteau [1]
Aurelie Larquemin [2]
Jyoti Prasad Mukhopadhyay[3]

2015 Open Data Research Symposium,
27th May 2015, Ottawa, Canada

Abstract

Public entities around the world are among the main producers of socio-economic data. The Open Government Data movement encourages and sometimes assists these entities to make their data publicly available, to improve transparency and accountability, leading to better governance. One of the paths is through research: OGD, by making available more and better data, should improve and multiply research studies and hence have an impact on evidence-based public policy. This paper discusses what the reality of the OGD initiative in India is for researchers and how OGD and its principles are considered and applied by the research community in India. We discuss the theory and the interactions observed between OGD and research in India, the advantages and challenges of using OGD in research and how to improve the use of OGD in research in India for more evidence-based public policies.

1. IFMR LEAD sharon.buteau@ifmr.ac.in
2. IFMR LEAD aurelie.larquemin@ifmr.ac.in
3. IFMR Business School jyoti.mukhopadhyay@ifmr.ac.in

1 Introduction

Through this study we aim to understand the usage, perception and opinion of researchers about the Open Data initiatives. Researchers in all areas rely on solid and quality data to conduct their studies and test their research hypotheses. One of the first steps in the development of a rigorous and systematic research study is to determine what kind of data is needed to answer the question(s) formulated and whether this type of data is currently available, and if not, then how to collect it. If the chosen topic of research is fairly specific then collection of primary data becomes inevitable for researchers. However, the cost of data collection, as well as the time and other resources required to do so often pose serious challenges. In that case secondary data appears to be the second best option available for the researchers, subject to its availability and quality.

Following the trend initiated by international organizations like the World Bank making cross-country GDP and related data easily accessible online, and the availability of open source statistical packages, statistics and econometric analysis of secondary data are increasingly accessible to all potential researchers and the number of empirical papers have increased substantially (Espinosa et al., 2012). These elements should lead to an improved quality of the research, as Open Data, by increasing transparency and replicability of the research studies, should reinforce scientific rigour. Data from different sources (e.g. government surveys, civil society organizations, industrial data repositories, private data producers, etc.) can be valuable sources of information. However, in the context of developing countries like India, this option remains relatively scarce. Open Data with quality checks in place in this case can be envisaged as a viable option to provide data and promote applied social science research.

The Open Data movement has encouraged and sometimes assisted various entities to make their data publicly available. For example with the Open Government Partnership, governments from more than 50 countries committed to improving Open Data access as a means for promoting major issues like economic growth, transparency, development, governance, etc. In recent years, the Government of India has undertaken a number of initiatives following the Right to Information (RTI) Act passed in 2005 and the elaboration of the National Data Sharing and Accessibility Policy (NDSAP): an Open Government Data (OGD) portal in India 'data.gov.in', National Data Bank, DevInfo India etc. There is also a movement advocating for making data from all publications publicly available and this movement is supported by international donors, multilateral funding agencies, think tanks, research centers, etc., and thereby these entities are encouraging the supply of data. The Abdul Latif Jameel Poverty Action Lab (J-PAL) and Harvard's Dataverse make the data used for a published research studies available in the public domain. Of late, several research funders have made the attribution of funding for research studies conditional on data dissemination to a wider audience.

Among the benefits of Open Data initiatives for research is that the publication of data helps avoid redundancy and therefore waste of resources. The publication of data following best practices and Open Data standards will help us to attenuate to a large extent duplication of activities like cleaning datasets, compiling, merging, formatting, etc. by dozens or hundreds of researchers. Thus, Open Data allows for a better allocation of research resources, which are often scarce. A database collected for one study can be used by another researcher to answer a completely different research questions. Hence, Open Data initiatives, by making secondary data available, could play a significant role in enriching the existing knowledge base at a negligible cost. However, not much is known about how researchers use this data and perceive the value of the Open Data movement. There is a dearth of studies that evaluate the impact of these initiatives on social science research itself, its quality, its extent, its translation into policy recommendation. Hence we propose to study these Open Data initiatives and their impact on applied social science research in India.

2 The Open Government Data in India, a work in progress

Access to quality Open Government Data which is essential for effective policy making in any developing country like India. Open Data is at the heart of Open Knowledge and transparency. Organizing Open Data Day on Feb 22nd every year across various countries has added further momentum to the Open Data movement across the globe. The idea behind this particular event is to promote adoption of Open Data policies by governments at all levels: national, regional and local[1]. Embracing the importance of data for good governance, the Government of India (GoI) has adopted a number of strategies. Next, we briefly discuss next some of these strategies and also briefly review the emergence of Open Government Data (OGD) in developing countries, including India.

2.1 Status

Open Government Data (OGD) refers to "data and information produced or commissioned by government or government controlled entities, which can be freely used, reused and redistributed by anyone"[2]. The first stepping stone to OGD is to set up a sound statistical system[3]. Towards this, GoI founded the National Statistical Commission in 2000. In 2002, the commission submitted a report with recommendations for strengthening and improving India's statistical system. This comprehensive report also identified the extent of critical data gaps in every ministry and government department. Based on the recommendations, the Indian Statistics Act 2008 and Collection of Statistics Rules 2011 were enacted. Around the same time, in 2005, India also enacted the Right of Information (RTI) Act which "mandates timely response to citizen requests for government information". Enactment of RTI was a commendable step towards greater transparency and good governance[4]. In fact, RTI and OGD can act in tandem to achieve the same goal: "to increase transparency of government by releasing information generated and collected by public funds for the citizenry to benefit from its social and economic value"[5]. In the same year, the GoI established the National Knowledge Commission, headed by Sam Pitroda, which made several recommendations to improve India's knowledge network. One such recommendation was to enhance government data dissemination through a national web-based portal for certain key sectors (e.g. Agriculture, Industry, Water, Energy, Environment, etc.). Later in 2006, the GoI also introduced the National E-Governance Plan (NeGP) with an overall goal of making government services more efficient, transparent, reliable and accessible to the common man of India through the Common Service Centre (CSC). This was a laudable move towards "good governance". In 2012 GoI went one step further towards active dissemination of government data by adopting The National Data Sharing and Accessibility Policy (NDSAP). Following NDSAP's inauguration, the OGD data portal 'data.gov.in' was launched later that same year.

2.2 Comparison with other developing countries

As per Open Government Working Group guidelines issued in 2007, OGD must be"complete, primary, timely, accessible, machine processable, non-discriminatory, nonproprietary, and license free"[6]. The launch of 'data.gov.' in the US in 2009 and 'data.gov.uk' in the UK in 2010 mark the beginning of the OGD movement. Subsequently, other developing nations started catching up and many developing nations joined the Open Government Partnership (OGP). For example, the launch of 'data.gov.ph' in 2014 marked the beginning of open data in the Philippines. This web portal features interactive dashboards, and infographics which support data visualization.

1. See http://opendataday.org/
2. Open Knowledge Foundation as cited in Yannoukakou and Araka, 2014, p.336.
3. See Srinivasan (2003) for a critical appraisal of the report submitted by National Statistical Commission (2002).
4. Under this act a person can obtain required information by submitting a request to the Public Information Officer (PIO) of the respective government department.
5. Access Info, 2011, p11 as cited in Yannoukakou and Araka, 2014, p.336
6. Yannoukakou and Araka, 2014, p.336

A similar portal 'opendevelopmentcambodia.net' exists for Cambodia. This portal "provides the public with up-to-date, accurate information about Cambodia and its economic and social development. Its open data approach guarantees materials and information are available to all users for use and download". The Peruvian counterpart of such initiatives is 'meetup.com/Hacks-Hackers-Lima' . Open data has also gained currency in recent years in Latin America. The first and second editions of regional conferences on Open Data (ConDatos) were organized in Uruguay and Mexico respectively[7]. In Ecuador, an Open data initiative, FLOK Society 'floksociety.org' is underway in collaboration with the Ecuadorian Ministry of Knowledge, the Secretary of Higher Education and the National Institute of Higher Learning. Moreover, enthusiastic researchers who are interested in exploring government budget and spending across different countries can log in to 'openspending.org.' In Africa, OGD is available at 'opendataforafrica.org', a program supported by The African Development Bank Group (AfDB). In South Africa, the Centre for Higher Education Transformation (CHET) has developed an online Open Data platform providing institutional-level data on South African higher education.

In 2004, the Reserve Bank of India (RBI), India's apex regulator of banking services, made its internal database on the Indian economy accessible to the general public. The RBI dataset called Database on Indian Economy (DBIE) is quite rich in terms of its contents[8]. The Open Knowledge Foundation in India has launched a web portal 'India City Open Data Census' to track the openness of seven major cities in India in terms of their Annual Budgets, Expenditures, Election results, etc[9]. Another civil society initiative worth mentioning is 'IndiaGoverns' which seeks to make development data accessible and useful for policy-makers, researchers, and the general public[10]. India Water Portal (IWP) is another such web based portal dedicated exclusively to water management knowledge dissemination[11].

3 Use of Open Government Data in research: a way to improve evidence-based policy making in India?

The common assumption about research is that the findings have a direct impact on the decisions of policy-makers and practitioners. We present in the first section the paths through which research can impact policy-making and how OGD can help reinforce this process. In a second section we assess the status of this process in India looking at the incentives and obstacles in the use of OGD by researchers.

3.1 How Open Government Data leads to better governance and evidence-based policies

The emergence of evidence based public policy making can be traced back to two centuries ago when a study was conducted in 1833 to show that education did not reduce crime (Andre-Michel Guerry's Essay on the Moral Statistics of France, 1883). In the 1970s it became evident that research findings most often failed to have an impact on policy-making. Investigations were conducted at that time to identify the failures in this process. What researchers learned was that their findings were only a minor component in the equation leading to policy-makers' actions. Other elements were, and are nowadays, still at play: political interests, ideological convictions, concerns about resources (staff and budget) to implement new activities, bureaucracy, weight of tradition, etc. However research findings were rarely acknowledged and most often drew very little attention from policy-makers. The use of research appeared indeed more complex than

7. For more information see http://condatos.org/
8. See http://dbie.rbi.org.in/DBIE/dbie.rbi?site=home
9. For more information see http://in-city.census.okfn.org
10. See http://www.indiagoverns.org/. Currently it focuses only on Karnataka.
11. See www.indiawaterportal.org

initially considered. This process was conceptualised and refined over the years. Whiteman in 1975 described a two-dimensional perspective on research use in policy making: a concrete and instrumental way for research findings to become a fundamental component of public policy and a conceptual and indirect path to influence policy making, by giving policy-makers a deeper understanding of issues in their field, new ideas or motivation and a new perspective on the targeted issues. In their work on research utilization in policymaking, Greenberg and Mandell (1991) adapted and refined this framework, considering that in both cases research could affect policy-makers in different ways, from a substantive manner to a more influential way or in a strategic purpose as described in Table 1. In this model each dimension, concrete and conceptual, is more envisaged as a continuum.

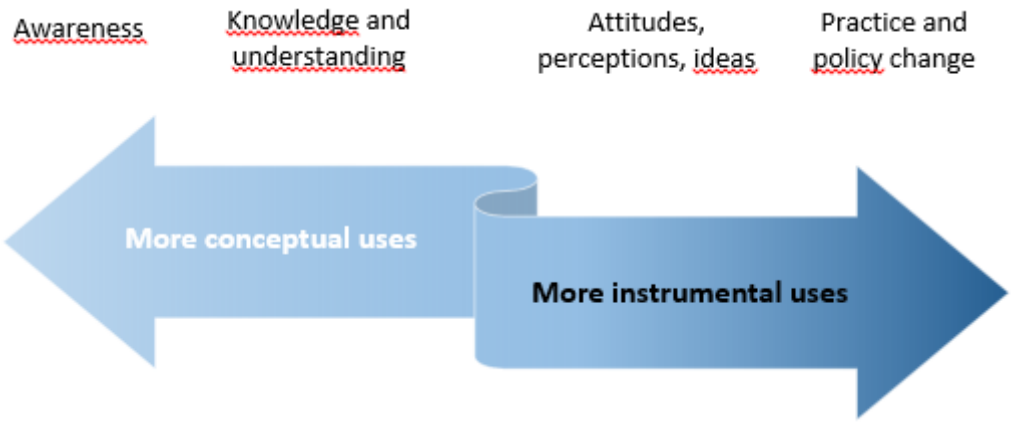
Table 1: Research use as a two-dimensional continuum

	Substantive	Elaborative	Strategic
Concrete	Research shapes the core of a decision or an issue	Peripheral use of research to further refine a position	Research is used to justify a position that has already been adopted
Conceptual	Research shapes a core orientation towards an issue or a basic understanding of the issue	Peripheral use of research to further refine an orientation or understanding	Research is used to confirm an orientation or an understanding that has already been adopted

Source: Adapted from Greenberg and Mandell (1991) and Nutley, Walter and Davies (2007)

This is also the retained framework by Nutley, Walter and Davies in their study on how research an inform public services (2007), who also argue that "evidence-informed or even evidence-aware policy would be a better description of the aspirations for the role of research in the policy making process".

Figure 1: A continuum of research use



Source: Adapted from Nutley, Walter and Davies (2007)

Several efforts have been put forth to identify this complex process and to eventually improve the use of research in public policy-making. David Webber, in his study "The distribution and use of policy knowledge in the policy process", explained that policy knowledge is not effective if it is not shared. It should be efficiently and extensively communicated and explained to policy-makers in order to impact policy decisions.

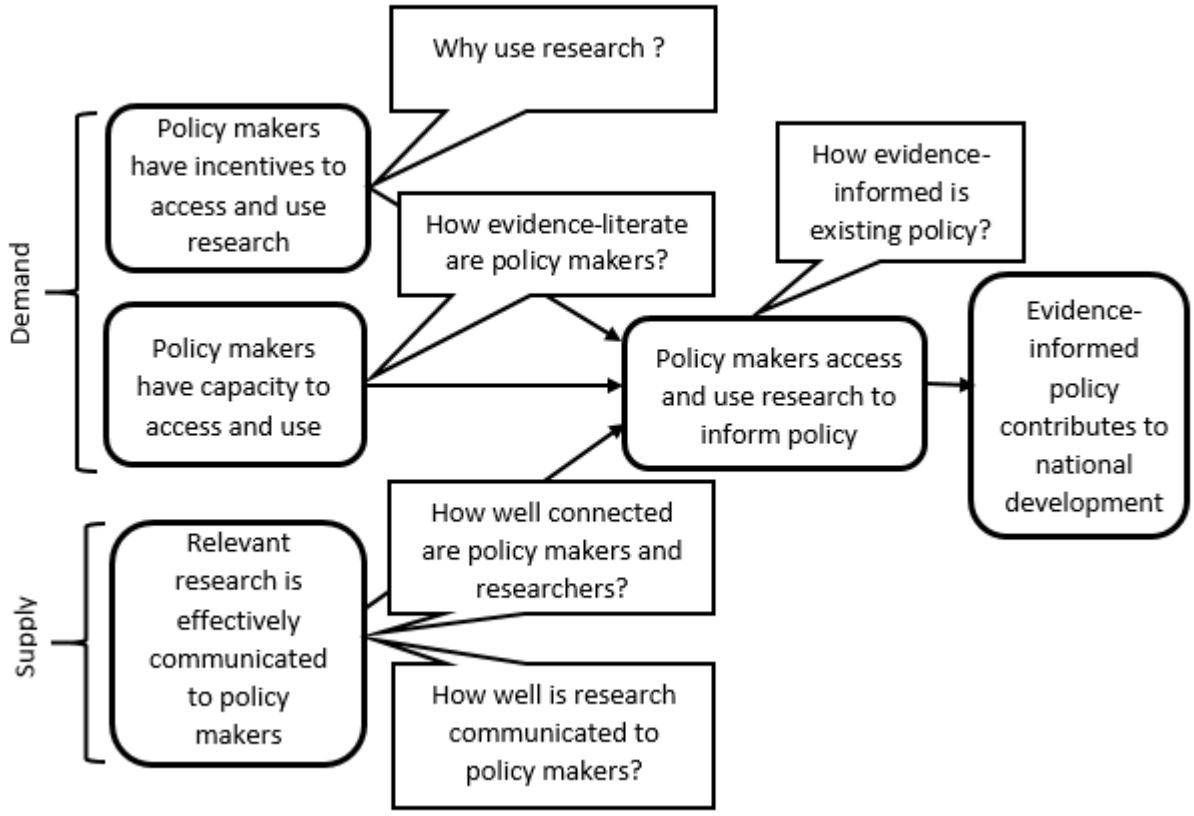
It is clear to the research community that the context in which the research is conducted has an influence on the uptake of the findings by policy-makers. Some conditions seem to predispose policy-makers to take research findings seriously. Developing countries have specific characteristics in policy making or in the ways to conduct and consider research, which can affect the uptake and adoption of research findings by policy-makers. According to Fred Carden in his study for IDRC (Knowledge to policy, making the most of development research, 2009) elements like the number of points of access through which research findings can flow, the openness of the system to the entry of new ideas, the democratic nature of decision making have a strong influence. He undertook an investigation to examine the consequences of 23 research projects funded by the IDRC and found evidence that development research, if done correctly, can improve public policy and accelerate development progress. His findings highlight the fact that research must be well designed and executed, and adequately and intensely disseminated to have the potential to influence public policy. However experience shows that development research frequently fails to register any apparent influence whatsoever in developing countries. Carden's study identified certain stylized facts about research and policy making environments in developing countries that could explain this failure. One such fact is that policy-makers often have less autonomy in developing countries. Staff turnover in research organizations and in government is high which weakens both research and policy influence. Developing countries often lack the intermediary institutions that may influence research to policy pathways. Implementation challenges are greater, both for research activities and policy. Researchers in development often lack hard data and may have to construct their own research-to-action machinery.

Other studies have led to similar conclusions: Court and Young (2003) undertook a comparative analysis of 50 case studies collected during the first phase of the Global Development Network (GDN)[12]. They have also identified gaps in the theory regarding the path of research to influence public policy due to the failure in accountting for specific characteristics of developing countries. According to his findings, "the key issue affecting uptake was whether research provided a solution to a problem. Policy influence was also affected by research relevance (in terms of topic and, as important, operational usefulness) and credibility (in terms of research approach and method of communication)". He also highlighted the importance of a clear and well-conceived communication strategy and strong advocacy efforts from the start, relating to the local context and concepts familiar to local policy-makers.

Following this evidence, international development agencies and other research funders are placing increased emphasis on the need to communicate research evidence to policy-makers. They not only consider the demand side of evidence from research, but also make sure that policy-makers have the incentives and capacity to access this evidence. On the supply side, it is recommended that researchers pay greater attention to their communication and dissemination strategy. Following the International Conference on Evidence-Informed Policy Making held in February 2012 in Ile-Ife Nigeria, Newman, Capillo et al. formulated a paper titled "What is the evidence on evidence-informed policy-making? Lessons from the International Conference on Evidence-Informed Policy-Making, INASP 2013)". In this paper Newman et al. present an updated model which explains how research results can affect policy making.

12. Bridging Research and Policy: Insights from 50 Case Studies, ODI, 2003.

Figure 2: Theory of change - Factors contributing to evidence-informed policy



Source: Newman, Capillo et al. (2012)

Focusing on this perspective, Lavis, Guindon et al. (2010) studied the engagement of researchers in efforts towards bridging the gap between research findings and policy making. They identified three sets of activities: providing systematic reviews of the research literature to their target audience, providing access to a searchable database of research products on their topic, and establishing or maintaining long-term partnerships related to their topic with representatives of the target audience. They surveyed 308 researchers who conducted research on one of four clinical areas relevant to the Millennium Development Goals (prevention of malaria, care of women seeking contraception, care of children with diarrhea, and care of patients with tuberculosis) in each of 10 low- and middle-income countries (China, Ghana, India, Iran, Kazakhstan, Laos, Mexico, Pakistan, Senegal and Tanzania). Their results show that even great research efforts remain poorly disseminated: less than half of the researchers surveyed reported that they engaged in one or more of the three promising bridging activities: 27% provided systematic reviews of the research literature to their target audiences, 40% provided access to a searchable database of research products on their topic, and 43% established or maintained long-term partnerships related to their topic with representatives of the target audience. Among the factors explaining the respondents' engagement in these activities were (i) the existence of structures and processes to link researchers and their target audiences; (ii) the stability in their contacts; (iii) and having managers and public (government) policy-makers among their target audiences.

OGD, by making available more and better data, should improve and multiply research studies and hence have an impact on evidence-based public policy. Requesting more Open Data could be part of researchers' strategies to bridge the identified gap between research knowledge and

policy. Open Data should ideally have the following characteristics: (i) availability and access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably downloadable over the internet. The data must also be available in a convenient and modifiable form; (ii) reuse and redistribution: the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets. The data must be machine-readable; (iii) universal participation: everyone must be able to use, reuse and redistribute (i.e. there should be no discrimination against fields of endeavour or against persons or groups) (Open Data Handbook, 2012). Many studies on OGD state that the release of OGD will improve the transparency and accountability of public authorities and lead to better governance. The publication of data increase government transparency and accountability, which generates confidence in government actions and encourage public engagement and willingness to participate in the policy-making process, leading to more efficient policies. According to Chun, Shulman et al. (2010), the OGD movement is part of the next step in governance ("Government 2.0"), with the release of data complemented by dialogue means and tools between citizens and public authorities. In May this year at the UCL-DIS Research Symposium: Open Data and Information, Prof. Shepherd in her address insisted that simply publishing data would not suffice to achieve the stated objectives. Additional work must be implemented to go from raw data to usable and reusable data with a supportive organisation which will ensure consistency and non-redundancy of efforts and guarantee data sharing, data integrity and quality. Each country is therefore entitled to build this requested system to guarantee the OGD to have the expected impact, including on evidence-based policies.

Public bodies are among the largest creators and collectors of data in many different domains. These data domains range from traffic, weather, geographic, tourist information, statistics, business, public sector budgeting and performance levels, to all kinds of data about policies and inspection (e.g. food, safety, education quality, etc.). The OGD movement in most cases assumes that public agencies are ready for an open governance process which considers influences, discourses and exchanges as constructive and welcomes opposing views and inputs. However, some studies have shown that in several countries the political class can be reluctant to adopt Open Government measures. Janssen et al. (2012) list the reasons for this reluctance which include the shift from a closed to an open system of governance, which has a significant impact upon relationships between the government and the civil society. Some remain sceptical regarding the expected positive outcome of the OGD initiatives compared to a system with barriers; so far no systematic research is available that addresses this concern.

Prof. Shepherd (2015) also pointed out that in countries where the OG movement has been initiated early on and effectively, there is strong support from the political class. She takes the example of the OGD movement's status in the UK pointing out that politicians argue that Open Data "would benefit the UK economy by creating jobs and stimulating innovation and at the same time increase transparency and accountability (...), empower the citizens' public participation and improve public services". She makes clear, however, that the return on investment for Open Data remains unclear, since "Open Data has no value in itself; it only becomes valuable when used". A multitude of studies regarding the potential economic gain following the implementation of OGD policies can be observed. In a report published in 2013, McKinsey Global Institute assumes that Open Data could have an important economic impact on societies. They announced an estimated figure of $3 trillion in annual economic potential that could be unlocked across seven domains: education, transportation, consumer products, electricity, oil and gas, health care, and consumer finance. These benefits include increased efficiency, development of new products and services, and consumer surplus (cost savings, convenience, better-quality products), without quantifying social benefits. They estimate the economic impact of improved education (higher wages), but not the benefits that society derives from well-educated citizens. They estimate that the potential value would be divided roughly between the United States ($1.1

trillion), Europe ($900 billion) and the rest of the world ($1.7 trillion). The study insists vigorously that, to attain a potential economic or political benefits, the OGD initiatives need to focus not only on release of data but also on the infrastructure, developers and researchers needed to make the data openness a reality and capture the full value of open data. Third parties, including researchers, must be involved and they should play an essential role in cataloguing, cleaning, analizing and other data related activities. Releasing metadata will make Open Data more usable. Investments in key skills including the ability to perform analyses, create useful reports and tools based on Open Data, and incorporate data into managerial decision-making processes is necessary.

There is consensus that OGD need researchers to maximise its potential in terms of better governance, accountability, economic benefits, etc. It is important to point out at this juncture that researchers also benefit from OGD initiatives. First, OGD offers opportunity to conduct new research studies by giving access to new datasets. Looking at these datasets they have the possibility to identify new questions and conceive new research studies. According to Guo Xu, (2012) among the benefits of Open Data initiatives on research is that the publication of data helps avoid redundancy and therefore waste of resources. The publication of data following best practices and Open Data standards allows dozens or hundreds of researchers to avoid going through the same procedures (cleaning datasets, compiling, merging, formatting, etc.). Open Data and OGD then allow for a better allocation of research resources, which are often scarce. A database collected for one study can feed another researcher's work on completely different research questions. Secondary data can be the key to important research work and the Open Data initiatives aim to make more secondary data available. Also Open Data in general and OGD, by increasing transparency and replicability of the research studies, should reinforce scientific rigour.

OGD and openness in research could stimulate research activities and outcomes, but this virtuous circle depends also on the willingness of researchers to share openly the data they use and their findings. On this issue, Azberger and al. (2004) perceive the openness of data as a way to ensure that both researchers and the public receive optimal returns on public investments in research. It also allows building on the value chain of investments in research, and its data resources with the underlying principle being that "publicly funded research data should be openly available to the maximum extent possible". Groves (2012) in his paper on open science and reproducible research also argues that data sharing can greatly increase dissemination, meta-analysis, and understanding of research results; it can also aid confirmation or refutation of research through replication, allow better implementation of research findings, and increase transparency about the quality and integrity of research. However, many have concerns regarding the open publication of their data and results. Prof. Shepherd rightly pointed out that researchers may be reluctant to release data, as there is a fear that original ideas and research will be stolen or misunderstood with the consequent danger of losing personal reputation. Some researchers also think that Open Data can be misleading or open to political manipulation. Other studies insist that the idea of open access to data in research is professed by many but not followed by much action (Hamermesh, 2006). The advantages to the research community and beyond generally fail to outweigh researchers' fears and costs to create the supply of data and replicable results (Anderson and al. 2008).

Versbach et al. (2013) used a dataset of 488 empirical economists from the top 100 economics departments and top 50 business schools to provide evidence of the status quo of data sharing or data access facilitation in economics. Using an ordered probit regression they identified that the likelihood to share is positively associated with sharing other material, being full professor, and being affiliated with a higher-ranked institutions. In another study Piworar (2011) studied 11,603 articles published between 2000 and 2009 that describe the creation of gene expression

microarray data. Using a multivariate regression, Piworar found that authors were most likely to share data if they had prior experience sharing or reusing data, if their study was published in an open access journal or in a journal with a relatively strong data sharing policy, or if the study was funded by a large number of NIH grants.

Our study focuses on the situation in India in terms of OGD. We intend to learn whether India is in the process of creating more and deeper academic research following the latest OGD initiative, to improve governance and the policy-making process.

3.2 Interactions between OGD and research in India: what we know so far

OGD is mostly released through the public entities' portals: the dedicated OGD portal 'data.gov.in' and also through an extended network of intermediaries like research centres and other NGOs. In this study we will focus on the OGD released at the national level, bearing in mind that public authorities of all Indian states release public data, which is also the case at lower levels (e.g. district or municipalities, etc.).

The OGD portal gathered datasets from the following government entities: Comptroller And Auditor General of India(CAG) (16); Department of Atomic Energy (1); Department of Space (15); Lok Sabha Secretariat (100); Ministry of Agriculture (367); Ministry of Chemicals and Fertilizers (14); Ministry of Civil Aviation (3); Ministry of Commerce and Industry (17); Ministry of Communications and Information Technology (18); Ministry of Corporate Affairs (26); Ministry of Defence (23); Ministry of Development of North Eastern Region (1); Ministry of Drinking Water and Sanitation (MDWS) (16); Ministry of Earth Sciences (10); Ministry of Environment and Forests (10); Ministry of Finance (124); Ministry of Health and Family Welfare (199); Ministry of Home Affairs (244); Ministry of Human Resource Development (67); Ministry of Information and Broadcasting (13); Ministry of MSMEs (14); Ministry of Mines (26); Ministry of New and Renewable Energy (12); Ministry of Panchayati Raj (3); Ministry of Petroleum and Natural Gas (26); Ministry of Power (5); Ministry of Road Transport and Highways (133); Ministry of Rural Development (2); Ministry of Science and Technology (92); Ministry of Statistics and Programme Implementation (345); Ministry of Tourism (3); Ministry of Water Resources (557); Planning Commission (776); Rajya Sabha (154).

This list covers a variety of cases and different degrees of openness of the data collected by public entities. We are considering in particular the status of three ministries in charge of collecting key datasets for socio-economic research: the Registrar General & Census Commissioner under the Ministry of Home Affairs, the Directorate General of Commercial Intelligence and Statistics (DGCIS) under the Ministry of Commerce and Industry, and the Ministry of Statistics and Planning Implementation (MOSPI) which conducts various surveys on Indian economic characteristics.

Indian Decennial Census

The Ministry of Home Affairs is in charge of conducting the Indian Census once every 10 years. The Indian Census is the largest single source of a variety of statistical information on different characteristics of the people of India. The Ministry publishes partial datasets by topic and geographic area on its website, the CENSUS DIGITAL LIBRARY (Beta-version, 54016 Census Tables from 1991 Census to 2011 Census (in Excel and Csv), and on the OGD portal. The Digital Library makes available all Census tables published from 1991 to 2011, Census Reports, and other digital files for free download in soft copy format.

The Ministry of Home Affairs has also initiated a Workstation for Research on Census Data at Jawaharlal Nehru University, Delhi. The workstation has been set up as part of a joint collaboration between the Office of the Registrar General & Census Commissioner and the

Jawaharlal Nehru University. It has the ability to access all the published census tables from 1991 to 2011 Censuses for research. Customized tables generated from the 2001 Census database can be made available on request from the researchers.The Census Digital Library does not collect any information on the data downloads, so it is not possible to estimate the use of the portal by researchers. A preliminary literature review does not show any significant increase in publications of research studies using the Census datasets.

India's trade statistics

The Directorate General of Commercial Intelligence and Statistics (DGCIS), Kolkata, under the Ministry of Commerce, Government of India, is the official organization for collection, compilation and dissemination of India's Trade Statistics and Commercial Information. It makes available the Provisional Monthly Foreign Trade data and the Final Monthly Foreign Trade Data. The Provisional Monthly Foreign Trade Data are posted monthly with a lag of around two months. The data for the current month and for the cumulative period beginning from April to current month of that year along with the corresponding data for the previous year are available on the old web-portal. For the Final Monthly Foreign Trade Data, the export and import trade data are available country-wise and economic region-wise. The detailed data on India's foreign trade in merchandise are usually available after three months from the reference month. The former website also makes available quick estimates for selected major commodities for exports and imports. The data dissemination policy of the DGCIS specified that a record may be any of the following single combination depending on the customers' specific requirement: a)Commodity*Country; b)Port*Commodity*Country, c)Commodity, d)Country, e)Port, etc. The payment toward the data cost was to be paid through Demand Draft in favour of 'Director General, DGCI&S, Kolkata', payable at any national or international bank at Kolkata.

DGCIS has improved the dissemination aspect with the creation of an online portal to disseminate authorized data. Trade data are available for download against payment of fees by credit/debit card online. No firm level/transaction-level data is available, but country-wise, port-wise, commodity-wise, data for the last 15 years as panel data is available. The portal provides for free metadata and more type of data is available: through this foreign trade data dissemination portal users can generate 11 different types of reports at 2/4/6/8-level of HS codes[13].

The list of datasets now available through this new portal is listed on the OGD portal, with a link to the page where one can access and purchase the data on the DGCIS data dissemination portal. In order to purchase the datasets the user must complete a form with his/her institution and designation. Information on users and use of data of this portal could then be extracted, but the DGCIS has not yet done this exercise, due to the recent launch of the portal and the regular modifications and additions that are still being made on a regular basis.

Ministry of Statistics and Planning Implementation (MOSPI)

Many datasets on socio-economic sectors are gathered by the Ministry of Statistics and Planning Implementation (MOSPI). The statistics wing, called the National Statistical Office (NSO), consists of the Central Statistical Office (CSO), the Computer Centre and the National Sample Survey Office (NSSO). The NSSO and CSO activities include compilation of National Accounts; conduct of Annual Survey of Industries and Economic Censuses, compilation of Index of Industrial Production, as well as Consumer Price Indices. The NSSO is in charge of the collection of socio-economic [14] , Urban Frame survey, Agriculture surveys, etc. The Ministry of Statis-

13. (a) Commodity-wise, (b) Country-wise, (c) Port-wise, (d) Commodity*Country-wise, (e) Country*Commodity-wise, (f) Commodity*Port-wise, (g) Port*Commodity-wise, (h) Country*Port-wise, (i) Port*Country-wise, (j) Commodity*Country*Port-wise, (k) Port*Commodity*Country-wise.
14. The Socio-Economic Surveys conducted by NSSO are in the form of Rounds, each round being normally of one-year duration conducted occasionally for a period of six months.

tics and Programme Implementation, Government of India, has been disseminating microdata from its surveys and censuses since long. This web-based survey cataloging system is powered by the National Data Archive (NADA) software developed by the International Household Survey Network (IHSN). The Microdata Archive provides web based access to the complete metadata of over 129 surveys and censuses conducted under the National Sample Survey (NSS) programme, Annual Survey of Industries (ASI) and the Economic Census currently available with the Ministry. Efforts are in progress to include metadata of surveys conducted by the State Governments and other government agencies. The metadata now provided in the archive includes, survey methodology, sampling procedures, questionnaires, instructions, survey reports, classifications, code directories, etc. Importantly, the archive provides all details of the data files that are disseminated to users along with case summaries. While the metadata is available, access requires a fee paid to the Computer Centre, which is in charge of the dissemination of the collected data by the Ministry. Hence, even if the metadata catalog is listed on the OGD portal 'data.gov.in' with the metadata accessible from this location, the database cannot be considered OGD. The process of payment is another obstacle to easy access of the data by researchers, as payments are to be made by demand draft issued by an Indian bank.

We have initiated the compilation of a list of articles regarding India on socio-economic topics from reputed specialised publications[15] and the datasets used in the papers, to identify if the datasets are in public access and from OGD. We intend to compare the trends of results before and after the implementation of OGD initiatives in India (Right to Information (RTI), National Data Sharing and Accessibility Policy (NDSAP), launch of OGD portal 'data.gov.in', etc.). Our preliminary results do not show any clear tendency towards greater access and use of OGD by researchers following the beginning of the OGD movement in India[16].

No study has been implemented so far to assess if obstacles are faced by researchers, both in India and abroad in accessing socio-economic databases collected by the government. It is important to assess whether the cost and payment method are the only issues encountered by the researchers. No study focuses on the Indian public authorities' dissemination policy and its rationale.

Chattapadhyay (2014) conducted a study to assess research and advocacy organisations' difficulties in access OGD pertaining to India. Among the challenges identified is the fact that most of the data collected by public authorities is not made available anywhere in digital format. The flows in the reporting structure between local, state and central public authorities prevent the publication of many datasets. Bureaucracy and reluctance to innovate also prevent the publication of original databases. These organisations also identify the public statistical machinery as out-of date, which impacts the data quality and the timeliness of data publication. There is a need to overcome these fault lines, to invest in the capacity of the Government agencies and reinforce their motivation to make the effort to publish original data, and to develop direct interactions between data producers and data users. This project also highlights the danger generated by the lack of publication of government data, which creates a space for a data reselling industry and a closed community of re-users of data. Public authorities are rarely aware of the existence of such 'a data black market'. Data intermediary organisations consulted for the study mentioned almost always downloading OGD from the government websites and almost never using the OGD portal, mostly because they were accustomed to collecting these

15. List of publications reviewed: Quarterly Journal of Economics, The American Economic Review, Journal of Applied Economics, and Journal of Asian Economics.
16. Summary statistics: From 2005 to 2015, no statistical evidence of increase in use of OGD or in publication about India. QJE (6 articles over the period, 5 before 2012 including 3 using OGD, and 1 from 2012 with OGD), AER (8 articles on India, 6 before 2012 with 4 using OGD, 2 from 2012, 1 with OGD), Journal of Applied Economics (22 articles on India, 10 before 2012 with 8 using OGD, 12 from 2012 with 10 using OGD), Journal of Asian Economics (6 articles on India, 4 before 2012, 2 from 2012, all using OGD).

datasets from the public authorities' websites long before the OGD platform was launched and they continued doing so. They also highlight that some OGD come most often in closed formats (that is, PDF files and HTML tables), even if recent efforts have been initiated to make datasets researcher-friendly. This poses an important barrier in converting data into usable formats.

In another study, Wright et al. (2010) identified the benefits and challenges of OGD in India. The first benefit identified is for the government itself: the data are more accessible for internal use as well, and its own information gathering and processing procedures can be improved as and when incorrect and outdated data are identified. It allows the GoI fulfilling its commitments towards the Right to Information (RTI). Among the challenges identified, an urgent issue is the entire infrastructure of information gathering, processing and sharing to upgrade, which is currently being implemented (e.g. New DGCIS portal). Insufficient standardisation is being denounced, as well as the need for system and semantic interoperability. Currently there is no use of common formats and software standards, and different departments gather different information under the same heading, or the same information under different headings.

4 Methodology of our study

Our study seeks to increase our understanding of the following: (1) Do researchers know about Open Government Data (OGD)? Do they use OGD? If so, how? if not, why? (2) What are the types and sources of open data used and why? (3) What is the perceived quality of OGD and quality of the research derived from OGD? (4) What are the issues of concern when using Open Data for research? (5) Are there any gaps in available OGD?

Research Design

Due to the nature of our institution and our activities and network, we have delimited our study to India and to socio-economic research.

To conduct this study we collected data and triangulated results from the following sources:
(1) In-depth desk review and study of the publication of papers focusing on socio-economic research and Open Data.
(2) Online survey focused on questions related to our study objectives. The survey was done among a purposeful sample from our extensive network of researchers, academics, and practitioners.
(3) Interviews with relevant stakeholders, and focus groups or roundtables were conducted to deepen our understanding and enrich our research findings.

The key findings from the literature survey have been presented earlier. They show a gap in terms of studies focusing on OGD and the impact on research, particularly in the socio-economic sector in comparison for instance with medical studies, and focused on developing countries. However we intend to verify some of the interesting elements that emerged from the literature survey, and also deepen our understanding of OGD availability in India through our survey exercise and meetings.

We are aware of the difficulty in gathering information from government representatives via an online survey from our previous experience. Hence, we focused our effort with this tool on researchers working on Indian socio-economic issues, based locally and abroad, and designed a detailed online survey questionnaire. We reached out to 19 of the top universities and schools in India actively engaged in research on socio-economic issues. We contacted both faculty members and current and past PhD candidates. We also shared this survey with researchers and PhD students based abroad in more than 70 distinct universities and schools.

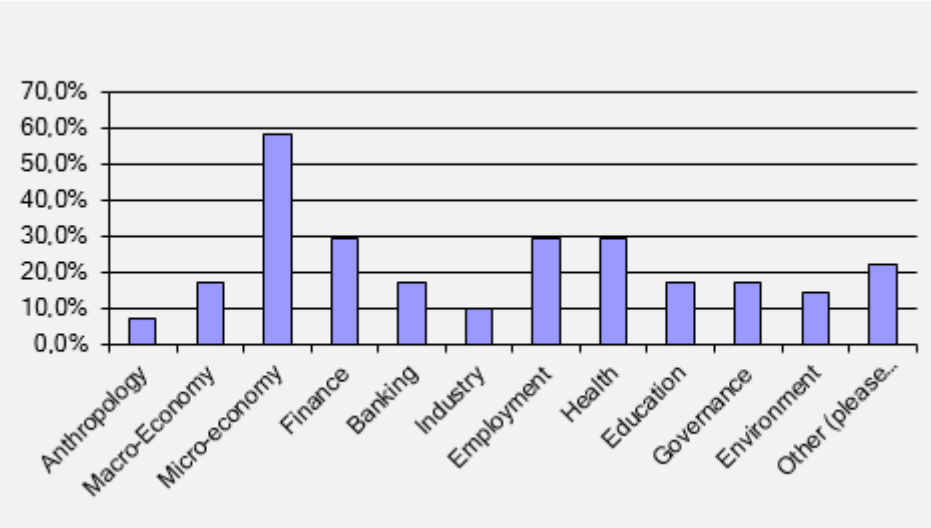
In the course of this study we interviewed members of the Ministry of Statistics and Planning Implementation (MOSPI) from the Computer Centre and from NSSO and CSO offices in Delhi and Kolkata. We conducted interviews with former members of NSSO and CSO offices, involved until recently in data collection exercises like the Annual Survey of Industry. We also organized a focus group, in the form of a one day roundtable or workshop in which several researchers and official statisticians participated. For the purpose of having a focused discussion we had to select a specific theme for this workshop, namely, Industry Data in India. Participants were then invited to discuss about Industry, Firms, MSME and employment data. Researchers, national and foreign, shared their difficulties working with such datasets, and offered suggestions for improvement and also cited examples from other countries. Official statisticians had the opportunity to share detailed insights about data collection and gathering procedures and their latest efforts to improve accessibility and quality of available data. A paper was prepared beforehand by two researchers to gather the views of the academic community and shared with official statisticians and other participants a few days before the meeting for their comments and feedback. All these elements led to important lessons presented in the following section.

5 Open Government Data and Researchers in India

5.1 Do researchers use Open Government Data?

Our online survey exercise received responses from 18 professors, 10 PhD or Postdoc Fellow and 36 research fellows and research practitioners which summed to 64 completed surveys. Our respondents were actively involved in research studies with 75% of them having been involved in a socio-economic research project over the last two years, in areas as diverse as anthropology, finance, environment, etc. Additionally 78% percent of them were involved in the design of the study, and 83% of them used secondary data for their research.

Figure 3: Domain of research of survey respondents

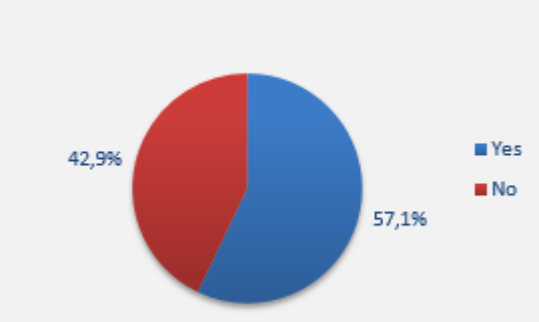


We designed our questionnaire based on our understanding from the literature review and the first discussions we had with the official statisticians, students and professors.

At the initial stage of our research we were under the impression that researchers were well aware of the OGD movement and such data availability in India, from either ministries' portals or ad hoc web-platform. From our survey responses we realized that OGD is still not very well-known

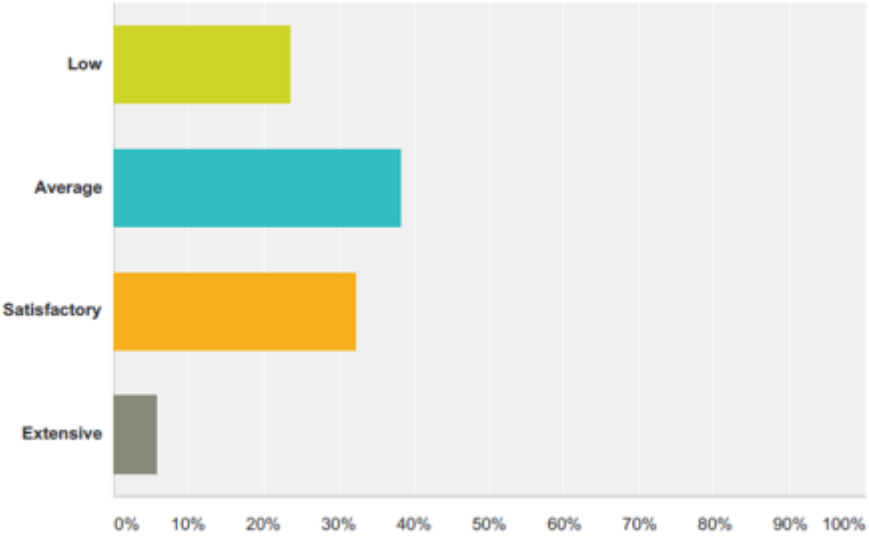
among the research community. Only 57% of our respondents declared knowing what OGD is, while a definition was provided in the survey introduction to briefly introduce them to Open Data.

Figure 4: Responses to the question: Have you heard of Open Government Data?



The degree of knowledge among researchers familiar with the OGD in India is quite low. Among those who responded positively to this question, 25% and 38% consider that they have a low or average understanding of OGD respectively, with only 32% assessing their knowledge as satisfactory and a small 6% judging it extensive.

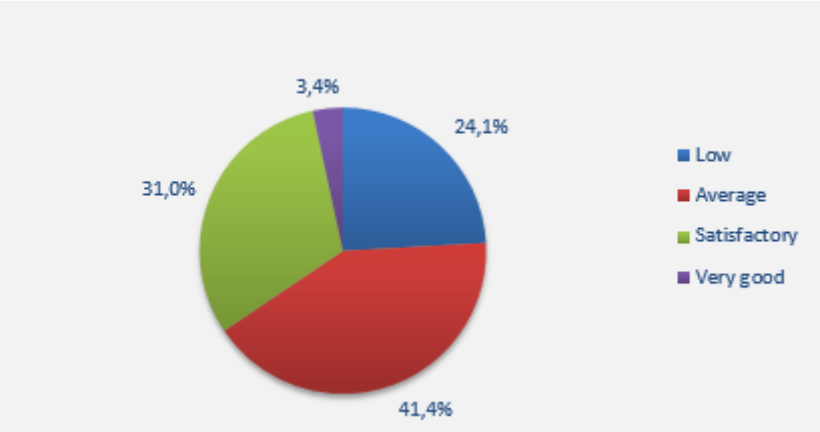
Figure 5: Responses to the question: How would you define your knowledge of Open Government Data sources?



Among the researchers familiar with OGD, only half of them declared having consulted the official OGD portal data.gov.in, while 88% of them have already consulted data from the various ministries' portal. These users appreciate mostly the easy access and the fact that such data was not available elsewhere to the best of their knowledge. However, many faced issues with downloading large quantities of data at once, and they regretted that the source and collection methodology were often unclear. Moreover, it was hard to get support from the official side, while some lament the lack of standardisation among datasets. A minority of the researchers who consulted these OGD sources did not work with those datasets eventually (18%). However, they were willing to work with OGD datasets in the future if some necessary improvements in the dissemination were made. Major importance is given in the provision of metadata, and

the possibility to reach out effectively for trouble shooting. The format of dissemination has to be standardised and be made more user-friendly. 65% of the respondents familiar with OGD judged their quality as average or below par.

Figure 6: Responses to the question: How would you rate the quality of Open Government Data available in the Indian context?



Of the respondents who had been involved in a socio-economic study in India in recent times, 80% used secondary data and they declared that they came from public sources, and not only from open sources. A large majority (60%) of them had to pay a fee to obtain the data (for a cost exceeding $100 for 70% of them). These datasets, free or not, were not accompanied by any metadata as reported by 70% of the respondents.

In terms of publication of their study data, only 58% of the researchers involved in a study with primary data collection were planning to make their dataset public. However among the researchers not planning to publish their data, 85% declared that they would do so if there was a portal gathering this type of datasets and research findings in their area of work in India. 90% were optimistic about OGD having a positive impact on the quality of research through better resource allocation thanks to less duplication of work (50% of the respondents) and due to a higher replicability of the studies (50%). But 60% of the respondents felt this can only be the case if there is an improvement in the quality of the OGD available. Open Government Data can have an impact on the public policy making process according to 98% of the respondents. OGD brings better involvement of officials (for 75% of respondents) and higher replicability of study and control by research community (for 54%).

5.2 Advantages and Challenges of using OGD in research

Among the main advantages listed during our study is the easy access of the OGD when they are available. The platform 'data.gov.in' is clear and allows users to quickly identify the database they seek. Many ministries and public institutions are also investing in making their platforms more user-friendly and the data more accessible. Another advantage frequently pointed out is the fact that the access to these OGD datasets is free. As previously mentioned, while exploring the OGD platform and other ministries websites, it was noticed that many key data for socio-economic research are only listed on these Open portals but remain accessible only after paying a fee. Data can be more or less accessible depending on the payment modalities (e.g. a demand draft issued by an Indian bank), as is the case for accessing MOSPI collected datasets.

The format of the OGD publication and dissemination has been unanimously criticised. Current formats are not consistent across datasets and sometimes across years, and require a fair amount

of researchers' time to make them amenable to analysis (e.g. when they are delivered only in hard copies on paper or in pdf format).

From the information we have gathered there is no concern about the intrinsic quality of OGD, or possible manipulation of data before release. However, researchers do express concern about the delayed release of data by public authorities, due to required preliminary treatments like anonimization, and to bureaucracy related issues and lack of good cooperation among the public institutions involved, leading to some datasets becoming outdated. For instance data from the Census 2011 have only recently been published, hence many studies conducted over the past several years had to rely on data from the previous Census of 2001.

A representative from the Computer Centre of MOSPI described the initiatives undertaken over the past few years to make their data more accessible and also more useful through the preparation and publication of metadata file. He explained that the OGD official website is used (by a satisfactory number of users per month) but is only a data repository where all official agencies only upload data without following-up on its usage. Moreover, this website generates many requests for information on the datasets, which are directed to the Computer Centre who cannot take charge of this aspect. The absence of troubleshooting tools and the lack of helpdesk features were also generally recognised as critical issues faced by researchers seeking to use OGD for their studies. He also admitted that the high turn-over within these services was inimical to the rapid implementation and success of these initiatives.

Another challenge among the data users we have interacted with is the absence of good quality or even total absence of metadata associated with OGD. A professor, specialist of sampling methods, talked to us about problems in sampling methods, which can be numerous due to the different existing methods to generate estimates for general populations from samples. The lack of information regarding the public entities' methodology for their data collection is an important issue. Understanding the available data is of utmost importance for researchers to be able to use data meaningfully. However a clear definition of the variables is rarely given and a variable name can have a different meaning depending on the data and the ministry considered, and may vary over time. A table of concordance sometimes exists within one entity, but not always. There is also the possibility that the data is conceived to be used at the national or state level but researchers may need the data disaggregated at a lower level (e.g. the district level, etc.), which requires many manipulations and compromising assumptions.

It is recognised that some efforts are being conducted to make more complete metadata available. MOSPI's datasets, for instance, are only listed but are not accessible through the OGD portal. On its own platform, 'the microdata catalog', MOSPI's data cannot be considered OGD due to the access charges. However, MOSPI is in a position to set a good example for other OGD portals in India. MOSPI has set up a metadata repository powered by the National Data Archive (NADA) software developed by the International Household Survey Network (IHSN).

While conducting interviews and round tables with researchers, many expressed their dissatisfaction with the low quantity of datasets published by the public authorities compared to the amount of data collected. This interferes in several ways with the research work: several professors from a top statistical research and training institute in India lamented the loss of power in their study due to the necessity of compiling different sources and datasets, with various adjustments and assumptions necessary to work around the absence of a certain dataset that the Government does not wish to publish. This is particularly true for micro-level data or data related to public finance.

For the roundtable dedicated to Industry, Firms, MSME and Employment Data in India, Asher and Novosad (2015) identified challenges faced by researchers in using these types of data. Most of them can be extrapolated from this specific area and remain very relevant to socio-economic data sources in India, including OGD. They noticed the difficulty in linking records among them, even when collected by the same public entity. There is also a strong need for uniformization and cooperation among official data producers. In the case of microdata at the firm level, for instance, a unique identifier could be put in place and be used through all datasets of this kind.

5.3 Suggestions to improve the role of OGD in research in India

Among the challenges complicating the interaction between OGD and researchers is the lack of awareness of the OGD resources available. In developing countries the OGD movement is less popular, in particular among junior local researchers. An awareness campaign targeting the research community could have great benefit for the use of OGD by researchers.

Regarding data collection by official statisticians, from our interactions it can be reported that they are aware of most of the limits identified here and are willing to make changes and improvements. MOSPI's top officers explain that a commission was in charge of improving the design and collection system in place, including reviewing suggestions made by the data users and researchers. They pointed out the limit of their resources and capacities, yet they recognise the necessity to make an important investment in such capacities and infrastructure to bring their data dissemination system to the next level. Officers reflected on how researchers could be involved in the reinforcement of these capacities, as they are also willing to participate fully in this effort. Suggestions included the development of internship programs to improve dissemination practices. Official statisticians have never denied the fact that many datasets remain unpublished, mainly due to this lack of necessary resources. Within public agencies and even within a single entity, it is crucial to improve cooperation and standardisation among data producers and data collection exercises.

Efforts and improvements are visible in terms of accessibility of data and provision of metadata. However, the OGD portal as it was conceived in India lacks the possibility to effect a real interactions among data users and data producers. This was a common conclusion from both sides: many official statisticians admitted that they did not know much about the use of OGD by researchers and that they did not have the structure or the resources to respond to the questions sent by researchers about the data.

To respond to the problems generated by inadequate data distribution formats, primarily the time-intensive work to prepare the data for use, datasets should be published in usable formats, like CSV files and must be accompanied by at least a brief description of the variables, data collection methodology and design, etc. Moreover, the data should be disseminated online, free of charge, to match the OGD definition. All public portals should employ a user registration process to gather useful information on data users and data use and should also include effective assistance tools and the ability to interact with data specialists. Due to the public resources scarcity, a third party of OGD resources, data intermediary organisation, could host a wiki or other documentation portal, and act as an intermediary between researchers with questions,and data producers and official statisticians with answers.

Asher and Novosad (2015) made a series of suggestions in their paper, applicable to all OGD. They suggest following best practices to use information technology in data collection. For instance, tablets and other electronic devices should be used for data collection, as it was the case with the 2011 Socio-Economic and Caste Census. This provides an immediate opportunity to improve data quality: time-stamps and geo-coordinates ensure that enumerators are actually moving from location to location and conducting surveys in a realistic time frame. They also

suggested setting up a data extraction centre, possibly through a third party (with secured machines, and export results tables, but unable to download confidential data.) as in other countries for confidential data. Such a data centre would provide access to required data after removal of confidential information and would be of great interest for researchers.

5.4 What about other Open Data resources?

This study focused on the OGD released at the national level and by public entities. There is a wider OGD network available in India with many public institutions at a lower stage publishing data as well as NGOs, think tanks, and other data intermediary organisations collecting and publishing data from public entities. Studying these oranizations provide a more precise picture of the OGD in India and its use by researchers because many of these organisations are conducting research activities. Valuable lessons could be learned from this research to improve the OGD portal and the way data is disseminated.
To improve the diffusion and use of OGD by researchers, it is imperative to pay attention, in future studies, to other data providers, such as private firms, and study their methods of data dissemination (in terms of format, metadata, modalities, assistance, etc.).

6 Conclusion

The OGD initiative is well implanted in India with a strong legal framework and support from public authorities. However, the movement remains unfamiliar to local researchers and could benefit from increased awareness and greater outreach efforts. Researchers aware of OGD generally used opened public data long before the launch of the specific OGD portal. If they value its accessibility they expect more from OGD: more data released, in a better and more user-friendly format, and data accompanied by relevant metadata. Those elements would help researchers to make better use of OGD in their work, and deserve to be prioritized by OGD organizations.

In terms of policy-making and the influence of OGD, the coordination and cooperation between researchers and public authorities must improve. Data users need access to assistance and to be able to ask questions on the portals where OGD are displayed. More efficient tools for the research community to report their needs in terms of data and for official statisticians to disseminate key information on the available data are necessary.

Public authorities are generally not aware of the research studies implemented using the data they produce. More interactions between data users and data practitioners is desired from both sides. It would lead to an easier and more efficient use of OGD in research studies, and the findings would be discussed with and better communicated to public authorities, which could influence public policies.

These stakes should be considered in the next steps of the OGD movement in India and possibly in other developing countries, and new tools should be designed to fill these gaps and support these challenges.

7 References

Access Info.org 2011, 'Beyond access: open government data & the right to (re)use public information'. Available from: http://www.accessinfo.org/documents/Access_Docs/Advancing/Beyond_Access_7_January_2011_web.pdf. [7 Jan 2011].

Anderson, RG, Greene, WH, McCullough, BD & Vinod, HD 2005. 'The Role of Data and Program Code Archives in the Future of Economic Research', Working Paper 2005-014C.

Andreoli-Verbasch, A & Mueller-Langer, F 2014, 'Open Access to Data: An Ideal Professed but not Practised', Research Policy, vol. 43, pp. 1621-1633.

Arzberger, P, Schroeder, P, Beaulieu, A, Bowker, G, Casey, K, Laaksonen, L, Moorman, D, Uhlir, P & Wouters, P, 'Promoting Access to Public Research Data for Scientific, Economic, and Social Development', Data Science Journal, vol. 3. Available from: http://www.ics.uci.edu/ gbowker/ promoting%20access.pdf [29 November 2004].

Asher, S & Novosad, P 2015, 'The Use of Firm Data for Development Research Around the World: Implications for India', International Growth Centre India Central.

Carden, F 2009, 'Knowledge to Policy: Making the most of development research', International Development Research Centre.

Chattapadhyay, S 2014, 'Opening Government Data through Mediation: Exploring the Roles, Practices and Strategies of Data Intermediary Organisations in India', Open Data Research Network. Available from: ¡http://www.opendataresearch.org/sites/default/files/publications/sumandro_oddc_project_report_0.pdf. [26 Oct 2014].

Chun, S, Shulman, S, Sandoval, R & Hovy, E 'Government 2.0: Making connections between citizens, data and governmen', vol. 15, issue 1,2, pp 1-9. Available from: http://citeseerx.ist.psu.edu / viewdoc/download?doi=10.1.1.257.6974_&rep=rep1&type=pdf [April 2010].

Court, J & Young, J 2003, 'Bridging Research and Policy: Insights from 50 Case Studies', Working Paper ODI 2003.

Davies, T & Alonso JM 2013, 'Researching the emerging impacts of open data in developing countries (ODDC): position paper', Prepared for the Open Data on the Web, Google Campus, Shoreditch, London. [23-24 April 2013]

Greenberg, DH & Mandell, MB 1991, 'Research utilization in policymaking: a tale of two series (of social experiments)', Journal of Policy Analysis and Management, 10 (4), 63356.

Hamermesh, DS 2006. 'Viewpoint: Replication in Economics', Canadian Journal of Economics, vol. 40, pp.715-33.

Manyika, J, Chui, M, Groves P, Farrell D, Kuiken, S & Doshi, E 2013, 'Open data: Unlocking innovation and performance with liquid information', McKinsey & Company 2013.

National Statistical Commission report 2002, Government of India.

Newman, K, Capillo, A, Famurewa, A, Nath, C & Siyanbola, W 2013, 'What is the evidence on evidence-informed policymaking? Lessons from the International Conference on Evidence-Informed Policy Making'. Available from: http://www.inasp.info/uploads/filer_public/2013/04/

22/ what_is_the_evidence_on_eipm.pdf [22 April 2013]

Nutley, SM, Walter, I & Davies, HTO 2007, 'Using Evidence: How Research Can Inform Public Services', The Policy Press, Bristol.

Open Data Handbook, Open Knowledge Foundation, November 2012.

Piwowar, HA 2011. 'Who Shares? Who Doesnt? Factors Associated with Openly Archiving Raw Research Data', PLOS ONE. Available from: http://journals.plos.org/plosone/article?id=10.1371 /journal.pone.0018657 . [13 July 2011].

Science as an open enterprise: open data for open science 2012.Available from: https://royalsociety .org/ /media/royal_society_content/policy/projects/sape/2012-06-20-saoe.pdf¿. [20 June 2012].

Srinivasan, T N 2003, 'Indias Statistical System: Critiquing the Report of the National Statistical Commission', Economic and Political Weekly, vol.38 no. 4, pp.322-37

Webber, DJ 1991, 'The distribution and use of policy knowledge in the policy process', Knowledge and Policy, Winter 199192, vol. 4, issue 4, pp 6-35. Available from: http://link.springer.com /article/10.1007%2FBF02692779 [1 December 1991].

Whiteman, D 1985, 'Reaffirming the importance of strategic use: A two-dimensional perspective on policy analysis in Congress.' Knowledge: Creation, Diffusion, Utilization 6: pp. 203-24.

Wright, G, Prakash, P, Abraham, S & Shah, N 2011, 'Open government data study: India', The Centre for Internet and Society. Available from: http://cis-india.org/openness/publications/open-government.pdf. [20 May 2011].

Yannoukakou, A & Araka, I 2014, 'Access to Government Information: Right to Information and Open Government Data Synergy', no. 147, pp. 332-340.