Working Paper

An investigation of National Open Government Data Platforms: How can India improve?

By Sridhar Ganapathy, Harsh Vardhan Pachisia, Rajeswari Parasa and Isalyne Gennaro

Data Governance Network

The Data Governance Network is developing a multi-disciplinary community of researchers tackling India's next policy frontiers: data-enabled policymaking and the digital economy. At DGN, we work to cultivate and communicate research stemming from diverse viewpoints on market regulation, information privacy and digital rights. We hope to generate balanced and networked perspectives on data governance — thereby helping governments make smart policy choices that advance the empowerment and protection of individuals in today's data-rich environment.

About IDFC Institute

IDFC Institute has been set up as a research-focused think/do tank to investigate the political, economic and spatial dimensions of India's ongoing transition from a low-income, state-led country to a prosperous market-based economy. We provide in-depth, actionable research and recommendations that are grounded in a contextual understanding of the political economy of execution. Our work rests on three pillars – 'State and the Citizen', 'Strengthening Institutions', and 'Urbanisation'. The State and the Citizen pillar covers the design and delivery of public goods, ranging from healthcare and infrastructure to a robust data protection regime. The Strengthening Institutions pillar focuses on improving the functioning and responsiveness of institutions. Finally, the Urbanisation pillar focuses on the historic transformation of India from a primarily rural to a largely urban country. All our research, papers, databases, and recommendations are in the public domain and freely accessible through www.idfcinstitute.org.

Disclaimer and Terms of Use

Suggested Citation

Ganapathy, S., Pachisia, H.V., Parasa, R., and Gennaro, I. (2021). An investigation of National Open Government Data Platforms: How can India improve? Working Paper 26, *Data Governance Network, Mumbai*.

An investigation of National Open Government Data Platforms: How can India improve?

December 2021

Abstract 4

Introduction 4

Research Question 4

Background and Literature Review 5

Evolution of Open Government Data in India 5

Data and Methodology 8

Data 8

Methodology 10

Technical Analysis 10

Policy and Regulatory Analysis 13

Findings 14

General Findings 14

Theme-wise Findings 17

Be demand Driven 17

Put data into context 18

Support conversation around data 19

Build capacity, skills and networks 20

Collaborate on data as a common resource 21

Analysis of India's OGD platform and Policy Recommendations 22

Technical Analysis of India's OGD platform 22

Policy Recommendations 23

Prioritise collecting and disseminating datasets that are in high demand 24

Join global open data partnerships and build regulatory frameworks 24

Develop a communications strategy to facilitate reuse of uploaded data 25

Scope for Future Research 26

References 27

Acknowledgements 30

About the Authors 31

Abstract

Open Government Data (OGD) platforms have recently become the channel of choice for governments to provide transparency and accountability. The purpose of these platforms is to enable citizens, the research community and the private sector to reuse public data for knowledge creation and innovation. OGD initiatives have been evaluated and benchmarked, and matured as platforms in the last decade. We study OGD platforms as infrastructure and evaluate them based on their capability, focusing on community engagement and supporting policies and frameworks that govern their creation and use. We find that the Indian OGD platform excels in many areas but highlights immediate and long-term changes that can help Indian OGD initiatives provide best-in-class engagement for users and champion a sustainable, healthy open data ecosystem.

Introduction

Open government data (OGD) has seen a global meteoric rise over the last twenty years based on a simple core idea- that government data should be a shared resource. The Organisation for Economic Co-operation and Development (OECD) defines such data as "a philosophy- and increasingly a set of policies - that promotes transparency, accountability and value creation by making government data available to all" (OECD, n.d.). Such a realisation has come about due to the massive economic value of open datasets that can be used to spur innovation when combined with privately-held datasets by enterprises. For instance, a 2011 survey conducted in the United Kingdom (UK) estimated that opening public data produced a social benefit of €8.3 billion (Government Valls I, 2014). On the other hand, there are other public benefits of publishing open government data, such as more inclusive service delivery, participatory decision-making and so on.

There are various examples of this such as helping control public spending by providing transparent budgets and expenses data (Cruz & Lazarow, 2021), assisting citizens to make informed decisions about their healthcare provider (Sangokoya et al., 2016), providing travellers with effective apps for a seamless public transit experience (Citymapper, 2018). However, all such use cases are only possible if OGD platforms are well designed and usable by various stakeholders. With that in mind, this paper aims to find out what Indian OGD platforms can learn from other countries to improve the design and usability of their platforms by third parties, i.e. individual researchers and innovators, private companies and civil society organisations, in creating new knowledge or services using such data.

Research Question

The goal of this study is to inform Indian policymakers such as the National Informatics Centre, Ministry of Electronics & Information Technology, NITI Aayog, etc. who are responsible for the development and maintenance of open data platforms put out by the government. In particular, we aim to showcase how these platforms can be used and redesigned by implementing best practices observed in other countries. We discuss technical requirements in light of the allied regulations and policies that countries/states must put in place. While the usefulness of such platforms to inform citizens and public discourse is acknowledged, we investigate and provide recommendations focused on aspects of knowledge creation and spearheading innovation. The research question is outlined as follows:

What can Indian open government data platforms learn from other countries to improve the design and eventual usage of their platforms by research communities (knowledge creation) and companies (for innovation)?

Background and Literature Review

Traditionally, countries provided access or the right to public information but not the right to reuse it. The Public Sector Information (PSI) Directive in Europe of 2003 and the open data initiative in the United States (US) in 2008 were the first steps in the direction of making government data publicly accessible. Open government data as a term arose in international debates in the early 2010s. In 2013, the G8 open data charter (ODC) set five international principles (International ODC, n.d.): "Open by Default", "Timely and Comprehensive", "Accessible and Usable", "Comparable and Interoperable", "For Improved Governance and Citizen Engagement", "For Inclusive Development and Innovation". The Open Government Partnership (OGP, n.d.) was also set up to help countries open their data through these principles. Since then, countries developed their own sets of principles and definitions. Today, open data platforms and initiatives are multiplying; the OGP counts 78 country members that commit to bi-annual plans to improve their open data ecosystems.

Evolution of Open Government Data in India

Efforts in India to openly publish data held within government departments was kickstarted by the Right to Information Act 2005. This was an initial attempt to improve the collection and dissemination of public data. The act aimed to "promote transparency and accountability in the working of every public authority, the constitution of a Central Information Commission and State Information Commissions…" (Right to Information Act, 2005). Section 4(2) of the Act states, "It shall be a constant endeavour of every public authority to take steps…to provide as much information to the public at regular intervals through various means of communications, including internet so that the public have minimum resort to the use of this Act to obtain

information." (Ibid.). The non-compliance with the requirement of proactive disclosure has been confirmed by research carried out by independent citizen groups. The report by Satark Nagrik Sangathan (2014) notes that, "nearly 70% of the RTI applications seek information that should have been proactively made public without citizens having to file an RTI application." Moreover, Wright et al., (n.d.) observe in their report that "one of the largest problems with complying with the proactive disclosure requirements of the RTI Act is that there is no easy system through which this data can be published online." This highlighted the need for a clearer framework, leading to the creation of the National Data Sharing and Accessibility Policy (NDSAP), 2012.

The policy, widely regarded as the genesis of open government data in India, aimed to "promote a technology-based culture of data management as well as data sharing and access." (NDSAP 2012). The policy applied to non-personal government data collected through various ministries, departments and bodies using public funds (Ibid). The implementation guidelines for NDSAP laid out the following principles for sharing data: "Openness, Flexibility, Transparency, Quality, Security and Machine readability." (Government of India 2015, p. 40).

Critically, the policy set up India's Open Government Data platform— data.gov.in. The platform was formally launched in 2012 and is managed by the National Informatics Centre (NIC) under the Ministry of Electronics and Information Technology of the Government of India. Following the launch of data.gov.in, five states have launched their own portals to host state-specific data and offer features similar to the national platform. Additionally, the Ministry of Housing and Urban Affairs (MoHUA) created SmartCities.data.gov.in, a portal for use by the 100 cities selected as part of the Smart Cities Mission, which provides city-level data published by the Urban Local Bodies. The India Urban Data Exchange provides a platform to facilitate sharing of data by cities to collaborate with industry, citizens and academia. Another initiative, Cityfinance is a portal created by MoHUA and supported by many civil society organisations to provide transparency about municipal finance in India. Each of these initiatives has seen varying degrees of success in the publication and reuse of data. Many cities have also been in the spotlight for leveraging their data capabilities to inform the response to COVID-19 (Deloitte & Ministry of Housing and Urban Affairs, 2020).

Literature Review on Open Government Data Platforms

As a result of the increase in global interest in developing and maintaining open data platforms, multiple frameworks and models have been proposed that should be followed when creating a data platform over the last decade. For example, Van der Waal et al., (2014) outlined the main functionalities of open data platforms while Yang et al., (2015) contrasted categorization structures by investigating the coherence, i.e. similarity, of the datasets in the same category. Solar et al.,(2012) proposed an open data maturity model that investigated the commitment of public bodies in pursuing the best practices of open data. In particular, the model followed a hierarchical structure consisting of domains, sub-domains and critical variables. Alexopoulos et al., (2014) extended this work by outlining a model of an open data platform that analysed variables related to the quality of the metadata, showcasing requirements for new datasets and explaining how best

to link disparate datasets. Lastly, Zuiderwijk and Janssen (2015) provided a set of six indicators to measure data quality for open data platforms while assessing the usability of participation mechanisms.

While the papers outlined above dealt with analysing open data platforms in general, platforms that contain datasets provided by government bodies must be analysed further. This is particularly important due to the massive number of stakeholders that utilize the platform from the suppliers (multiple government bodies providing data and a central agency managing the platform) to the end-users ranging from researchers, private companies, citizens, etc. Maali et al., (2010) conducted one of the first studies evaluating OGD platforms by analysing platforms from five countries to identify patterns and overlap in their structure. Barbosa et al., (2014) conducted one of the largest assessments of data platforms by analysing over 9000 platforms from 20 cities in the United States. Oliveira et al., (2016) built on this work by applying it to Brazilian OGD platforms. Umbrich et al., (2015) evaluated 82 active open data platforms across 35 different countries.

Ubaldi (2013) states it is difficult to find a standard mechanism for the evaluation of OGD initiatives. Hence, they propose an analytical framework that assesses OGD platforms based on legal & implementation aspects and the socio-economic value of the datasets that are released. On the other hand, Kalampokis et al., (2011) analysed 24 OGD initiatives and provided a classification scheme based solely on the technology utilized to publish the datasets. Braunschweig et al., (2012) also presented a survey of existing OGD platforms, focusing on their technical aspects. Petychakis et al., (2014) then analysed the OGD sources developed in the EU27 from a functional, semantic and technical perspective, in terms of their thematic content, licensing, multilingualism, data acquisition, data discovery, data provision and data formats.

Sayogo et al., (2014) also analysed platforms from 35 countries and provided a framework for evaluating OGD initiatives such as usability, software engineering or standards compliance. Verma and Gupta (2012) compared 30 national-level data platforms based on data release formats and found that the development of OGD platforms follows an incremental approach, similar to those of other e-government initiatives. Most of these papers evaluating OGD platforms looked at features of such platforms including retrievability, usage, completeness, accuracy, openness, availability of metadata, standardisation, discoverability and machine-readability of data. Ultimately, Charalabidis et al., (2014) developed a methodology that aimed to evaluate OGD platforms based on the estimation of value models from users' ratings. The authors found that policymakers should give the most priority to improving data upload and the search and download ability of datasets. This is because these features seemed to be the most important based on user feedback while also having the highest impact in terms of economic value generation.

In the Indian context, Máchová and Lněnička (2017) evaluate the performance of the Indian OGD platforms, together with other countries, using select criteria and find that it ranks in the top five. In the latter category, Saxena and Janssen (2017), use the Unified Theory of Acceptance and Use of Technology (UTAUT) to study the uptake of OGDs among citizens. They made use of survey

findings from 244 respondents who were students, teachers, and bureaucrats in Indian cities. The study found that the use and acceptance of India's national data platform (data.gov.in) did improve among the sampled population. This could, in part, be due to having better resources such as steady internet connectivity, better awareness due to the kind of social networks they belong to. On the other hand, when Buteau et al., (2015) focused specifically on the experience of researchers concerning the Indian OGD platforms, they found that there was little awareness of OGD among local researchers. Further, public authorities entrusted with producing and maintaining datasets did not have information about the kinds of research being produced with these data. Hence, while access and awareness about OGD might have improved, there is a lot to improve in India's OGD ecosystem.

Data and Methodology

We took a two-pronged approach to evaluate the data platforms, both from a technical and policy standpoint. As a first step, the dataset of the platforms to be analysed were identified and selected as outlined below.

Data

Since the launch of the first open data platforms by the US in 2009 and the UK in 2010, an increasing number of countries have launched similar open data initiatives and data platforms to make it easy for the public to find and use these data- they are available in a range of different formats and span through a wide range of domains. As observed by Umbrich (2015), the number of datasets and platforms seem to be continuously growing over time.

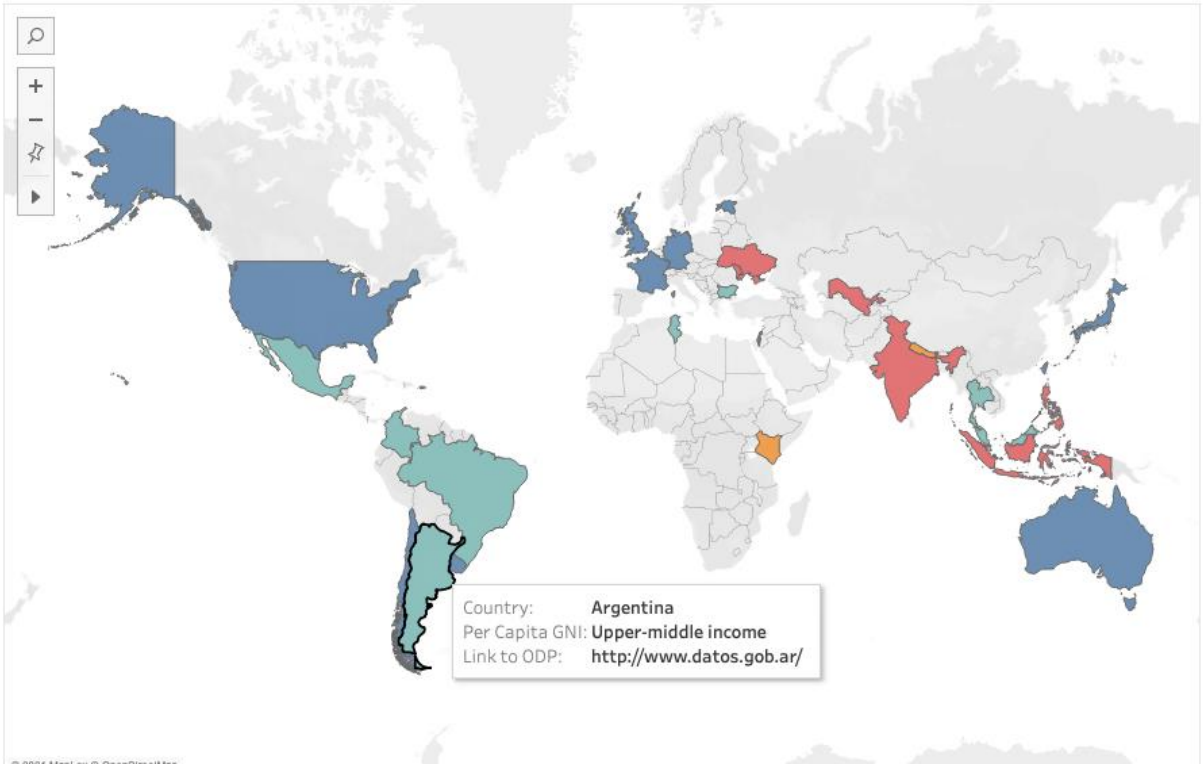
For our study, only open data platforms at the national level are evaluated, no international, regional or local open data platforms are considered as well as national statistical institutes or non-official platforms, which may also offer open data. The base dataset was generated via lists of OGD platforms outlined by the Open Government initiative by the United States as well as Máchová and Lněnička (2017) for a total of 78 data platforms. To narrow down this dataset and ensure a holistic approach, countries listed in the platform were chosen based on per capita GNI provided by estimates from the United Nations (Figure 1), as well as geographical location (Figure 2). By applying these filters, the list was narrowed down to 33 platforms.

Figure 1: Countries based on UN per capita GNI

Income Categories	Number of Countries analysed
High Income	12
Upper-middle income	8
Lower-middle income	6
Low income	2

Source: UN Per capita GNI estimates

Figure 2: Geographical distribution of Open data platforms analysed



Source: Based on the author's estimates

Next, the verification of the platform's existence consisted of the following steps. First, the name of the country chosen was inputted into a general search engine alongside the keywords "open data", "government data" or "open data portal" or "open data platform". Second, the identified platform's URL was opened to examine whether it was in working condition. Third, a check was made to ensure that the platform chosen was developed and maintained by a government agency by examining the 'About' section of the website itself. Following this verification, only 28 of the platforms were accessible as of November 2020, which were all analysed based on the methodology outlined below.

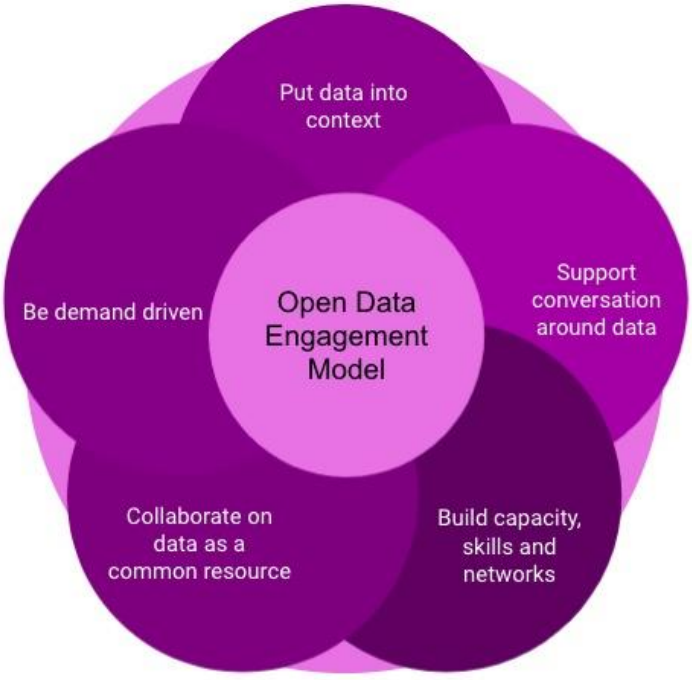
Methodology

We took a two-pronged approach to analyse the selected platforms. First, we developed an evaluation framework to compare each platform against the other based on technical parameters. This framework was created based on a systematic review of past work done in this field and complemented from our own experience and requirements as members of the policy and research communities in India. Second, we undertook a deep dive into the regulatory and policy evolution of open data platforms in a few select countries to supplement our technical analysis and provide actionable recommendations for Indian policymakers.

Technical Analysis

We use the themes of the '5 stars of Open Data Engagement Model' developed by Davies (2012) as a base layer to structure our analysis (Figure 3). This model, developed specifically for evaluating open government data engagement, is a non-technical parallel to Berners-Lee (2010) well-documented 5 stars linked data model. Using these themes as a base to analyse OGD platforms, we chose 12 quantitative variables outlined by Máchová and Lněnička (2017) and Sethi and Prakash (2018) to evaluate each of the platforms. These authors had developed frameworks that looked at both the general characteristics of data platforms and datasets themselves. For this paper, only platform-specific characteristics were utilized and placed within the thematic framework outlined by Davies (2012).

Figure 3: Five stars of Open Data Engagement Model



Source: Davies (2012)

The themes and the variables include the following:

1. Be demand-driven

This theme explores whether the datasets, tools, and support that are released by the government authority are structured and based on the needs of their end-users (i.e. researchers, private companies, etc.). Moreover, it aims to test whether each platform has developed an adequate feedback mechanism to listen and respond to requests from its community.

To achieve these aims, we chose two quantitative variables namely 'platform usage statistics' and 'Application Programming Interface (API) availability'. Platform usage statistics was chosen because transparency about the platform's contents and performance is paramount for establishing trust with users. Moreover, it helps users to understand which publishers are prolific in their publishing cadence (upload and update frequency), and which datasets are considered high-value by the community (view and download counts). Next, API availability was investigated since providing programmatic access to data is the most reliable and sustainable way to ensure that access is end-user agnostic (Grzenda & Legierski, 2021).

2. Put Data into context

This theme explores whether government authorities provide clear and transparent information about the datasets uploaded. In particular, it tests whether details about the frequency of updates, the formats, how it was created and the quality of data is available. It also evaluates whether metadata is provided and if manuals have been provided to assist end-users to make use of the datasets themselves. Lastly, it investigates whether knowledge creation by end-users is assisted by providing links from the data catalogue pages to prior analysis or tools developed by the community.

To answer these questions, we evaluated platforms based on the variables of 'reuse tracking', 'similar datasets list' and 'instant visualisation'. Reuse tracking was included since it is an effective mechanism to showcase the value and assist end-users to build on existing work, rather than start from scratch. We also checked whether a list of "similar datasets" was included alongside any dataset. Each dataset can build on another and hence, suggesting alternatives and potentially relevant matches for user searches makes dataset discovery easier in the longer term. Lastly, instant visualisation was included since it lowers the barrier to data usage by acting as a preview of the data, primary check of data quality, and helps users understand the nature of the dataset.

3. Support conversations around data

This theme questions whether end users can comment on the datasets or have structured discussions with the community on the datasets uploaded. In particular, it evaluates the ease of use in joining such conversations, connecting end-users to the supplier of the dataset and whether

it promotes further offline discussions of the use of the datasets available. To answer such questions, we investigated whether the platforms had the following: a social media handle, the ability to share datasets from the platform, and if a forum exists to facilitate a feedback mechanism as well as conversations around the data. As we all know, social media has become critical in reaching out to audiences and hence, having in-built linkages to the popular social media channels helps improve the reach of the platform by enabling engagement across end-users. Still, a forum also remains crucial since it creates a feedback loop by providing the ability to engage with data publishers or researchers on a common platform and promotes further innovation and knowledge sharing.

4. Build capacity, skills and networks

This theme explores whether public authorities provide linkages to other tools that would assist end-users in working with the datasets provided. Moreover, it questions whether guidance via training sessions/videos was provided on how to use analysis tools to assist end-users in making the best use of the datasets on the platform.

To answer these questions, we analysed the usefulness of the information provided on the 'Help' and 'Frequently Asked Questions (FAQ)' sections of the data platform's website. The Help section is important since it can range from questions and answers about data availability to detailed PDF/video guides and API usage. Such resources are crucial to grow the user base and enable proactive stakeholders to engage with the platform. On the other hand, the FAQ section can contain essential information including but not limited to the legislation that has served as the basis for the OGD initiative, the type of data that the authority intends to share on the platform, licenses governing the data on the platform, and contact information for feedback. Both can also provide details of training sessions on using the platform if they are provided by the government authorities.

5. Collaborate on data as a common resource

This theme cuts to the main aim of opening government data: that it should be a shared resource. With that in mind, it questions whether policymakers are collaborating with end-users to create new datasets that are derived from older ones and based on conversations. Next, it asks if the platform provides support to help end-users build and sustain tools and services that directly work and connect with the data uploaded.

To evaluate this theme, we looked to see whether data could be requested and how much metadata was available. A tool that provides the ability to request data is crucial for two reasons. One, it helps gauge interest and provides information to data publishers about which datasets they should make publicly available. This is done by providing users with the ability to request and vote on data that is yet to be made public. Second, it has an indirect impact of making the community feel heard, ensuring that end-users return to the platform. Having a dedicated channel to accept data requests makes it easier for users to provide their input and for data

publishers to respond to such requests. By providing detailed metadata, policymakers allow users to evaluate the availability and quality of the published data. Doing so gives the community a clearer view of the platform's worth as a resource in the longer term.

Policy and Regulatory Analysis

While all 28 platforms were analysed from a technical standpoint, we undertook a deep dive into a few policies and laws that pertain to the development and implementation of OGD platforms in those countries. Current regulations and policy frameworks for France, Germany, Brazil and Tunisia were taken from our sample and analysed. These countries were selected since:

1. Each of their OGDP performs relatively well as per our technical analysis.
2. The countries are at different levels of economic development.
3. All four are members of the Open Government Partnership and Brazil and Germany are federal systems, like India.

In particular, we identified the set of policies and legal provisions that seem to enable the *"proactive release of large volumes of information in formats and under conditions that permit re-use"* through such platforms, especially by third parties such as research organizations and so on.

We highlight specific interventions, interesting policies and discuss the regulatory framework that guides the development of OGD platforms in these countries. Then, we summarize these learnings through a quick framework where we associate such provisions with the objectives of community engagement that such platforms must achieve. In all, the following steps were undertaken:

- Identification of general trends and patterns on policies related to the development of the platform.
- Investigation of the regulatory framework by finding public authority responsible and reviewing main policy documents.
- Highlighting the following:
    - Use cases and examples of policies entailed towards improving the open data platform from launch till the date of analysis.
    - Examples of research communities using it for knowledge creation.
    - Examples of private sector/start-ups using it as inputs for innovation.
    - Examples of governments creating systems of incentives to encourage reuse and engagement of the research/innovators community.
- Providing recommendations to improve the Indian open government data platform and the country's OGD initiative based on the insights gained.

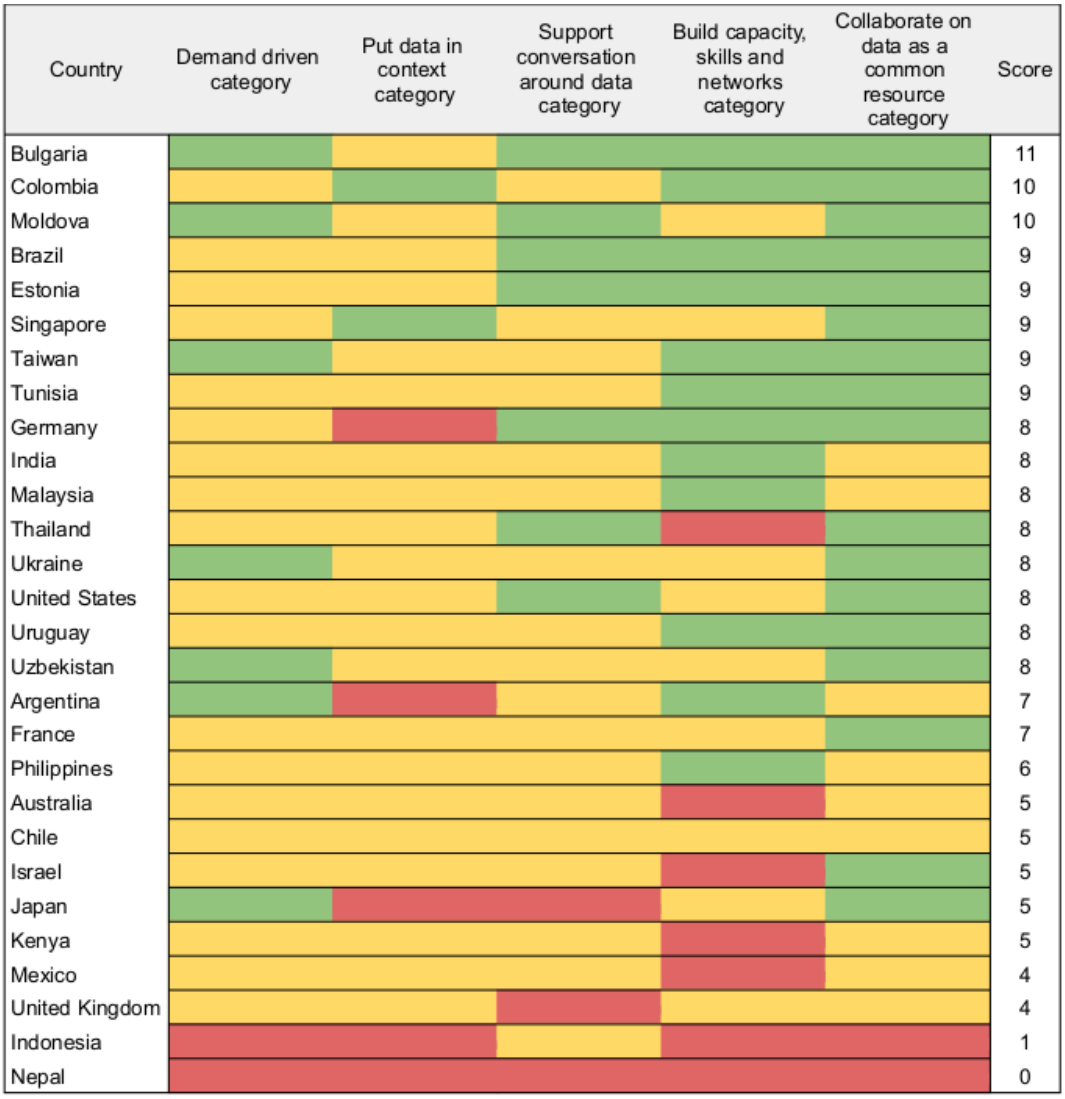
The findings of this technical and policy analysis of the platforms investigated are provided in the following section.

Findings

General Findings

60% of platforms in the sample excelled in the 'Collaborate on data as a common resource' theme whereas only 7% did so in the 'Put data in context' theme, offering insights into the overall strengths and general focus of the platforms. All platforms use a similar thematic grouping of datasets to provide users with an easy way to discover and explore data. This helps showcase data of closely linked departments together. For example, the commerce section has data about foreign trade, industry, and corporate affairs. 50% of the platforms made their OGD initiatives open-source, allowing the community to contribute to the platform, and offering them guidance and technical resources needed to host their open data platforms. The United States' platform was the oldest in the sample, established in 2009, and the newest was from Indonesia and Nepal, both of which were launched in 2019. A general overview of how each platform scored based on our technical analysis is outlined in Figure 4 below.

Figure 4: Analysis of national open data platforms



Source: Based on authors' estimates

From our policy analysis of the open government data policies and initiatives taken by four countries (Brazil, France, Tunisia and Germany), we found that each of them created an open data friendly ecosystem by formulating strategies and involving the community at an operational level. To do so, they did the following:

- Identified high-value datasets and focused on specific sectors
- Created various incentives to build and keep the open data community-engaged, from hackathons to national strategies
- Developed action-oriented policies, with concrete projects that respect fundamental principles of open government data, and clear licensing
- In the longer term, they all created corresponding legislation to develop the sector with legal provisions to incrementally improve the sector

The key learnings and highlights from each of the platforms are outlined in Figure 5 below.

Figure 5: Key learnings/highlights from France, Germany, Tunisia and Brazil

Country	Member / Signatory in the international community	Authority responsible for open data sector or formulating policies at the national level	Highlights
France	G8 Charter for open data OGP member	Etalab	• Focuses on mandatory "reference data" • Builds and fosters the community • Focuses on encouraging innovation within public administration, and collaborations with the private sector for developing apps etc. • Has worked on developing open data legislation for improved reuse of government data.
Germany	G8 Charter for open data OGP member	Federal Ministry of the Interior, Building and Community	• Encourages innovation through its national High-Tech strategy 2025. • Identifies key and additional datasets • Has adequate legislation in place
Tunisia	Open Data Charter OGP member		• Works with different sectors through working groups to identify key datasets. • Has a communication strategy for the administration and the citizens • Aligns all strategies and action plans in one programme: multi-pronged strategy with concrete projects and clear targets, aligned with OGP action plan, is comprehensive, identifies clear governance structure for implementation. • Assesses of status quo of digital and open administration, openly • Defines "reuse" and "public data" very clearly. • Complies with international standards. • Involves the private sector and the civil society at all policy-making stages, from the strategy to the operation as a key factor of success.
Brazil	OGP member		• promotes the participation of the non-government organizations or research institutes at the planning stage to work on socio-economic purposes later • Enforces implementation across levels of government • Create incentives for innovation and proactive reuse of data at all levels - from Federal to city-level

Source: Based on the author's estimates

Theme-wise Findings

The overall findings (categorized by theme) based on the technical analysis of the platforms and buttressed by relevant policy initiatives are outlined below.

1. *Be demand Driven*

10 platforms (35%) have dashboards displaying downloads/views of datasets each month and allow visitors to discover the most popular datasets, most used datasets, or publishers with the most datasets. Website analytics has been identified as the preferred and most popular way to measure data use by national statistical offices (Sethi & Prakash, 2018) and widespread adoption by OGD platforms is necessary for platform providers to understand their audience and be demand-focussed. The French platform has an exemplar usage tracker that provides granular insights on a wide range of website engagement aspects including approximate geographic locations of users visiting the platform; the number of returning and unique visits; time spent on each page of the site; average duration of each user session on the site; a dataset-wise number of unique and total downloads; stats on the referral pages through which users land on the open data platform; and more. Such detailed engagement metrics can help understand user behaviour on the platform and cater to their interests. Platforms of Estonia, Ukraine, and Moldova insightfully present the platform usage statistics as well. All three of them present data about downloads over time that can be viewed for each publisher.

24 platforms (86%) offer API access to datasets and 19 of those offer unrestricted access (no prior registration or user-specific API keys). Such tools help the user access public data anonymously. To make them more accessible, guides should accompany API keys. Such documentation should include a quick start guide, endpoint definitions, code snippets, and example responses to the requests made along with links to other useful and relevant resources. Colombia's platform is a great example of this providing such documentation alongside authentication information and details on limits on requests with and without an API key (the latter requiring a sign-up).

Providing an option to explore data based on dataset usage gives users and publishers insight into demand patterns. Taiwan allows sorting by the number of views and downloads, whereas France uses the number of times that data was 'reused' and the number of 'subscribers' to the dataset as metrics. This is a crucial way for OGD initiatives to move from purely making data available and focusing on the audience and need for data (Sieber & Johnson, 2015).

Identifying and publishing high-value datasets that meet the demand of citizens, researchers and the private sector is also critical. This creates scope for innovation and is crucial to promote the

reuse of open government data. We found that Germany, France and Tunisia focused on high-value datasets to publish in the public domain. France mandates them under its law, whereas Germany and Tunisia identified them in their action plan.

The Federal Ministry of the Interior (2014) details Germany's national plan to implement the G8 Open Data Charter. It identifies key and additional datasets to be published. The key datasets are the federal budget, geodata, elections and federal statistics. The additional datasets cover a broad range of sectors, from the foreign office, a digital library for culture and media, a register of agricultural enterprises, the natural forest reserves or the police crime statistics. Additionally, the Federal Chancellery (2019), in its OGP Second National Action Plan for 2019-2021, claims to be active in the ongoing European process for identifying high-quality datasets which will be available free of charge, in machine-readable formats and via APIs, across the European Union. France, on the other hand, mandates the publication of "reference data" (Government of France, 2021). Reference data are defined as datasets frequently used by people - other than the proprietary authorities - to identify products, services, territories or people.

As can be seen from the above examples, each country defines high-value datasets differently. In France, for example, these datasets are considered basic public information that is frequently used by people for identifying other variables. On the other hand, in Tunisia, the sectors of focus chosen were ones that are central to building an open government and could align with the overall development strategy of the country. Ultimately, choosing and spending resources defining the right, high-value datasets is key for a developing country like India. Doing so at its early stage of open government will help the country overcome current capacity constraints. Moreover, it will help avoid wasting resources publishing high volumes of data across sectors, while compromising on data quality.

2. *Put data into context*

14 platforms (50%) share details of data reuse on their platform. Only eight of these provide users with meaningful navigation of the reuse sections (allowing users to filter APIs, apps, infographics, services, etc.). Taiwan and France's platforms offer a good model for showcasing reuse and linking it to the data used.

14 platforms (50%) allow users to preview and visualise data without requiring a download. Having a wide range of visualisation types is an effective tool to allow users to instantly engage with the data. Colombia's platform stood out the most as it offers a dashboard view with up to nine chart types along with data filtering capabilities. Every dataset also has "live" connectors allowing integrations with external tools like Tableau, Plotly, Excel Power BI, Carto, and so on. While some platforms allow viewing data tables along with basic functionality like sorting, some others also offer graphs and map views at the dataset level, as in the case of Tunisia, Australia, Moldova, and Uruguay.

Many of the platforms analysed have specific features that assist in inputting the data uploaded into context. These include:

- Germany's platform uses a map search feature that lets users select an area on a map and view data that is connected to those areas.
- Tunisia has a dedicated space on its platform for developers to access data through APIs. They can also monitor the consultation of their reuse through statistics on the consultation, creating their profile and contributing to the forum.
- Through mobile apps, online services, visualisations and articles showcased on its platform, Tunisia encourages researchers, journalists, officials, civil society and citizens to share their data reuse cases.

In the long term, to adequately support conversation around government data, it is critical to developing legislation that allows for its reuse. For instance, France has well-developed legislation that regulates the reuse of open government data (Government of France, 2021). Within this code, article L. 321 details the reasons behind which datasets cannot be reused for other purposes. The exchange of public information between authorities for public services isn't considered as reuse. Moreover, if adequate rights of intellectual property are cited, the authorities are not allowed to oppose the reuse of the data. Further, the general rules (Ibid.) establish the code of conduct towards the reuse of public information as follows:

- It is subject to conditions of not altering or denaturing them, and their source and data of latest updates must be mentioned.
- The proprietary authority is in charge of the reuse of the data if it is possible only after anonymisation of personal data.
- The absence of response of administration for the reuse of public information is considered as a refusal
- The administrations producing or holding public information must keep a repository of the main documents in which the information is located at the disposal of the users. This repository gets updated every year. It must mention the exact title, the content, the date of creation, the conditions of reuse and the date and purpose of the updates for each document. The authority must publish it on its website.
- The administration producing the information must comply with open standards to publish the conditions for reuse and/or fees and rationales for fixing fees.

Going forward, India should adopt both the features highlighted across different platforms to make datasets uploaded more accessible to a wider audience and promote the sustainable reuse of such data through long-term regulation.

3. *Support conversation around data*

17 platforms (60%) feature their social media handles on the website and 15 (53%) provide social media sharing links on each dataset's page. Many platforms use the reuse section as a way to keep their visitors updated about open data meets and hackathons.

15 platforms (53%) allow users to add their comments for each dataset or provide a dedicated forum for users to discuss data. Australia's approach stands out as it provides a dedicated forum with standard engagement metrics for each discussion (number of views, replies, and users in the conversation). Github (used by 2 platforms), Facebook, StackExchange are some of the other tools platforms use to engage with users. France's platform helps further the conversation by showcasing datasets added by the community, relevant reuses, and past discussions for each dataset.

Through our policy analysis of specific countries, we found that it is critically important to have a communication strategy that familiarises the administrations internally and the public at large with the idea of open data. If people know what to expect from open government data, they will keep the government accountable and prepare the administration for the task of dissemination. For instance, Brazil planned to build knowledge across levels of government in the process of designing its Open Data Policy (Ministry of Transparency and Comptroller General of Brazil, 2018). It also defined a few concrete milestones to measure the implementation. In particular, the strategy to create the open data system included mapping the existing processes for disclosing data, organising a public consultation to discuss the draft model, and establishing a communication plan for three levels of government and civil society.

Tunisia, on the other hand, designed a communication strategy as part of its governance framework to implement its national programme SmartGov2020. The team was in charge of developing a campaign to augment the visibility of the programme in the administrations and to sensitise the public about the socio-economic dimension of the projects. Ultimately, making conversations surrounding the uploaded datasets easier through such strategies and tools would catapult the development of India's open data ecosystem.

4. *Build capacity, skills and networks*

17 platforms (61%) have a Help section that educates users about the platform's offerings. Brazil's platform is available via the Brazilian Sign Language through their Vlibras (Government of Brazil, n.d.) initiative. Israel's platform helps visitors with vision impairment or colour blindness.

16 platforms (57%) have an FAQ page to address potential issues faced by users. Brazil's platform uses a Wiki to share all resources related to open data publishing and provides a public view of the decision-making process used to improve the platform. 36 per cent of platforms provide users access in multiple languages, but most are inaccessible to a global audience without Google Translate.

The USA's Open Govt is a good example of policy translating into action on this front. The Data Act (2019) requires the government to provide information about tools, best practices, and schema standards in open data. This helped create a comprehensive collection of resources (US Government, 2019) for skill/capacity building, ranging from case studies about developing a

system to providing seamless access to a database of all federal regulations (Federal Register, 2010).

For the success of an open data platform, supply-side constraints on uploading datasets are also essential to be tackled when building capacity. To this end, the Tunisian Association of Public Auditors held training sessions with 45 municipalities on open data and developed an open data platform for urban local bodies through the allied "Onshor" programme. In France, the Mission "DATA" (Government of France, 2020) under TechGouv aims to build financially and technically capable public administrations to manage data-based public policies. It develops skills in data science to analyse data for framing policies and in data visualization and prospective for measuring the impact of policies that are implemented. Finally, to aid capacity building, countries have also built partnerships with each other to create a collaborative network. For example, the open government partnership ensures an equal level of engagement of the civil society, private sector and government, which helps improve the state, build capacity and keep the government accountable to its objective. The partnerships also facilitate cross-learning between nations (South-South Facility, 2012). As the Indian open data ecosystem develops, updating national strategies and addressing similar supply-side constraints is essential to building long term capacity.

5. *Collaborate on data as a common resource*

22 platforms (78% of the sample) offer users ways to request data, of which 17 platforms have dedicated forms/pages where users can request, with 5 others providing an email ID or forum where users may post their suggestions. Taiwan's and Colombia's platforms display data requests publicly along with comments by other users and actions taken in response to the request. Giving users a complete view of how the platform handles requests builds engagement and trust.

21 platforms (75%) allow users access to the platform's metadata directly or through APIs. This also has the additional benefit of allowing initiatives like Portal Watch to assess data quality on the platform or for researchers to provide feedback about crucial missing elements in the platform's data offerings. It also increases the reach of the platform by allowing other aggregators to automatically harvest and host the data, for instance, Data Portal Asia (2021). Estonia encourages (OGD Estonia, n.d.) users to contribute public data to the platform from sources that may not be available on the platform. France provides tools (OGD France, n.d.) for users to automatically add data from a user's website to the OGD platform.

To further collaborate on using data as a common resource, some governments create incentives for public innovation through hosting events. For instance, Tunisia organises a national hackathon, the OpenGovDataHack to bolster the reuse of open data (OGP Tunisia, 2020). Civil society projects from such a hackathon ranged from mapping the spatial distribution of health workers to management of the national airline TUNISAIR flight delays using the data provided. On the other hand, Brazil has organised multiple hackathons, but at a city-level with similar results. Successful products from them include Cidadão Recifense (Recife Citizen) and Rio

Inteligente (Smart Rio) to improve the delivery of healthcare services for citizens and tourists, with a "health-unit finder based on user location", and "an online vaccination card." Both the apps emerged from the cities' app development contests in July and August 2013.

The inherent question of licensing also becomes critical when talking about collaborating on data as a common resource. Severe restrictions placed upon reuse via stringent licensing inhibit the extent of usability by researchers and private companies. Each country analysed had different methods to resolve this issue. These are highlighted below:

- Germany's act on reuse of Public Sector Information (PSI) regulates conditions of reuse; the federal government will comply with the Open Data and Public Sector Informative Directive of the European Union by 2021 (The Federal Chancellery, 2019).
- In Brazil, Oliveira et al., (2016) found that, until 2016, 35.79% of the datasets on the federal government's platform did not have a license description. They estimated that, across a large sample of all the Brazilian OGD platforms available at the time, almost 71% of datasets are licensed under the Open Data Commons Open Database License (ODbL), edited by the Open Knowledge Foundation.
- In Tunisia, the license is established under the Tunisian legal system and conceived to be compatible with all free licenses such as OGL of the UK, CC-BY 2.0 of Creative Commons and the ODC-BY of the Open Knowledge Foundation.

The licence by the Open Knowledge Foundation for commercial and non-commercial reuse of public data. Public data includes all datasets that are/must be published or at the public disposal and that is provided or collected by all public agencies according to their service delivery mandates. Under this framework, the data publisher, i.e. public authorities, is responsible to ensure that the data is not the intellectual property of another party. The publisher is also obliged to the Open Data Charter principles. The license authorises personal rights, non-exclusive and free to reuse the data. This is available worldwide and for unlimited time. By learning from such measures, Indian government data can truly be used as a common resource.

Analysis of India's OGD platform and Policy Recommendations

Technical Analysis of India's OGD platform

There are a variety of good practices that the Indian OGD platform does undertake. For instance, it offers data via APIs, has a social media presence to connect with the community, and has seen a consistent increase in the number of datasets on the platform each year (Darpan-National Informatics Centre, n.d.). It also possesses inbuilt data visualization tools — a feature present in half of the analysed platforms — to explore datasets allowing users to create their posts that can be shared with the community. It also contains a data suggestion feature (available in 60% of platforms) that allows the public to view each suggestion received and its popularity via up. Till August 2020, this feature has received 2600 unique suggestions since 2012 and only 18 of those

have over 100 votes. In terms of information, the Indian platform's sector page provides a broad overview of the resources available for easy navigation and provides links to data, blogs, visualisations, forms, and documents. Ultimately, on our index, the Indian platform scores 8 out of 12 in the indicators (along with 7 other platforms) and is slightly above average in the sample. 12 platforms (43%) rank below India, and 8 platforms (29%) are ranked higher.

While there are positives to the Indian OGD platform, there is a lot of scope for improvement. In particular, the platform needs to improve the following features: machine-readable formats, make information on authors available for all the datasets, ensure open licensing and copyrights, etc.

At present, India's platform does not have a well-maintained general discussion forum (present in 40% of platforms analysed) wherein interested parties can exchange ideas or participate in a conversation about the platform's datasets and resources. Once developed, it can be further built into a dedicated channel for receiving feedback by engaging directly with the user community. The platform also lacks a detailed usage report (present in 35% of platforms) that shows which datasets are in demand by the community. Going forward, the platform needs to implement the objective dataset quality rating using the five-star data model, as referenced in the NDSAP. Doing so would provide information about data quality and tangible steps for the improvement of dataset quality.

In the spirit of interoperability and openness, the platform should also promote API-based harvesting of metadata from the platform or offer periodic metadata extracts. An example of this is the metadata provided by the OGD platform of the United Kingdom. This encourages independent evaluation of the existence and quality of datasets in the platform. Additionally, it helps avoid the formation of disconnected information dumps (Jasinskaja, 2019) and created federated data platforms to access the Indian platform's catalogue, thereby offering potential re-users additional paths to search and access data.

Next, while the platform does contain a data visualization feature, it can be made a lot more seamless if users are allowed to create visualizations from datasets directly alongside instant loading of data onto the dashboard. In the same vein, barriers need to be reduced to allow API access to data, which at present requires an account creation and key. Further, while India's platform offers an API sandbox to allow users to run and test requests, this can be supplemented with more human-readable information on the fields or columns that will be returned from the requests, quick steps on getting an API key, and query options along with links to any further helpful resources. Ultimately, by implementing such technical design changes to their platform, the Indian OGD platform would massively improve and result in cross-cutting knowledge exchange and innovation.

Policy Recommendations

Alongside revamping the Indian platform from a technical perspective, longer-term policy changes need to be undertaken to structurally improve the Indian OGD platform and the OGD

ecosystem in the country in general. At present, open government data is a relatively nascent stage in India. This can be observed since, by the end of November 2020, 53 of the 106 ministries and departments did not have an active Chief Data Officer. Hence, the government can learn from what has worked and hasn't in more advanced countries as highlighted through this paper and in the specific policy recommendations below.

1. *Prioritise collecting and disseminating datasets that are in high demand*

We recommend that India focuses on developing and using a platform at the national level, sharing data internally, improving data collection and management practices, and making authorities accountable to certain datasets. Similar to France, the government can adopt the principle of "reference data", i.e. datasets that are mandatory for producing and releasing in publicly available format and can be reused. This could be using data (OGD India, 2017) belonging to high-value information that has seen demand from past community engagements (OGD India, 2016). The federal structure of India's administration implies that all levels of government will build thematic and customised open data platforms. Each platform needs to comply with established principles of interoperability (Drafting Team Metadata & European Commission Joint Research Centre, 2007) and not work in silos within the open data ecosystem (UN Statistics, 2018).

Engaging with future consumers of open government data from the start helps design adequate laws and platforms and shows the government the value that its data hold for innovation and development. This paper defines such consumers as groups of researchers, businesses and civil society organisations that are likely to track the datasets and keep the government accountable for opening them. Their sectoral expertise and experience doing research, working with citizens and creating goods and services can help the government understand the standards and designs of platforms that enable reuse. Experts can focus on the standards that will ensure good quality for publishing, i.e. granularity and frequency of publication. Additionally, open data experts can inform principles of Open Data policies and legislations. For example, the Department of Science & Technology recently solicited public comments and reviews on the draft National Geospatial Policy (2021) which aims at improving the Geospatial Sector in India to acquire and produce geospatial data, maps and allied products and services. Similarly, for the open data sector, the government can draft new policies based on consultation of core experts and request public reviews to refine them.

The government can also establish multidisciplinary groups of experts, similar to how Tunisia proceeded, and focus on priority sectors for innovation or access to information. The groups of experts can identify which are these essential datasets through surveys in their industry and research. Also, when realising the potential of its data the government will trust opening up its data.

2. *Join global open data partnerships and build regulatory frameworks*

Making too precise laws, too soon, can be restrictive and inappropriate. To avoid this, it is important to build knowledge, and capacity first. For this, the Indian government should consider re-joining the Open Government Partnership which has helped other developing countries in this sample, such as Bulgaria (OGP Bulgaria, 2011) and Colombia (OGP Colombia, 2011), to perform well (see Figure 4). Action plans need to have measurable objectives too. The first axis of Tunisia's 2018-2020 action plan (OGP Tunisia, 2018), for example, is "enhancing the right of access to information and opening up public data". This is broken down into various commitments, each commitment is measured based on explicit expected results. Among other things, the government of Tunisia expects to educate civil servants and raise awareness of officials in public institutions on this front. Also, the reporting and review of the partnership is a great process to keep India accountable for its engagement in building open data platforms and related legislation. It tracks the achievement of the government's commitment and facilitates cross-learning between members. This approach enables innovation and empowers administrations to use their judgement and knowledge that the legal frameworks may otherwise confine.

To back up the next actions that India will take, we recommend making incremental changes to the existing regulatory framework, rather than trying to define new mandates and legal provisions with great accuracy. Brazil, Germany, France and Tunisia did not have all the mandates in place when they started developing the open data sector in the early 2010s. France's legislation has evolved from the laws on electronic communication established in the 1980s, to very detailed mandates and the adoption of today's European framework (GDCR). Brazil did a few concurrent iterations on its policies and platforms since 2011 too. It enforced its OGD platforms policy after 2 years. Initially, it focused on developing OGD platforms at all levels of governments, across the country without standards which limited the potential for reuse of the data and inconsistencies on the Federal government platform which had to be revamped later on. The Indian government can build upon the relevant policies dealing with data privacy, interoperability, publishing open data and access to information such as the NDSAP, the Government Open Data License (2017), the draft Data Protection Bill, the IT act and e-government policies. Institutionalising best practices and promoting existing standards in incremental policymaking can help enforce data sharing and transparency laws for government agencies and set a stepping stone to regulate open data in the country.

3. *Develop a communications strategy to facilitate reuse of uploaded data*

A communication strategy is a key to understanding the value of open government data and design platforms that facilitate the reuse accordingly. It should aim at informing citizens about the responsibilities of the government to open public data, and of informing all authorities about the benefits of doing so. A communication strategy would include national and sub-national hackathons such as Germany and Tunisia, training of administrative staff, as the Mission DATA in France (Government of France, 2020), and opening and bringing awareness for citizens to keep the government accountable. It is essential to increase the user base of open data. Proactive release of negative lists and data publishing schedules like the Ministry of Statistics and Programme Implementation (MoSPI, 2019) provides prospective users with clarity of

expectations from the authority and predictability in the data publication cycle, which is essential for community engagement.

Scope for Future Research

While this paper showcased how Indian open government data platforms can be used and redesigned by implementing best practices observed in other countries, there is scope for further research on the topics that are not analysed.

First, delving into the quality and collection processes of individual datasets that make up the entire open data platforms explored in this paper is necessary. Answering those questions would require an entirely different study with access to internal government officials in charge of collecting and uploading data onto the relevant platforms. Second, research should be conducted on the larger political economy and governance constraints that inhibit a policymaker's capability to share data collected with the larger public.

Third, further analysis of the open data ecosystem at sub-national levels is critical since such platforms have the potential to directly impact the lives of citizens given their granular reach. Fourth, from a technical perspective, device-specific differences (computer/mobile) in a user's experience of the platform should be explored given the exponential rise of smartphones across the country. Finally, further research on the aspects of data governance and privacy in regards to open data should be discussed. All these topics would benefit from further research as they all have an impact on the role that open government data can play in society.

References

Alexopoulos, C., Zuiderwijk, A., Charapabidis, Y., Loukis, E., & Janssen, M. (2014, September). Designing a second generation of open data platforms: Integrating open data and social media. In *International Conference on Electronic Government* (pp. 230-241). Springer, Berlin, Heidelberg

Barbosa, L., Pham, K., Silva, C., Vieira, M. R., & Freire, J. (2014). Structured open urban data: understanding the landscape. *Big data*, 2(3), 144-154.

Berners-Lee, T. (2010). *Linked Data - Design Issues*. W3. https://www.w3.org/DesignIssues/LinkedData.html

Braunschweig, K., Eberius, J., Thiele, M., & Lehner, W. (2012). The state of open data. Limits of current open data platforms.

Buteau, S., Larquemin, A., & Mukhopadhyay, J. P. (2015, May). Open Data and Applied Socio-economic Research in India: An Overview. In *Open Data Research Symposium*.

Charalabidis, Y., Loukis, E., & Alexopoulos, C. (2014, January). Evaluating second generation open government data infrastructures using value models. In *2014 47th Hawaii International Conference on System Sciences* (pp. 2114-2126). IEEE.

Chief Data Officers. (n.d.). Open Government Data Platform India. https://data.gov.in/datacontrollers

Citizen Recifense. (n.d.). http://www.cidadaorecifense.com.br/

Cityfinance.in. (n.d.). http://www.cityfinance.in/home

Citymapper. (2018, May 30). *Building a city without open data - Citymapper*. Medium. https://medium.com/citymapper/building-a-city-without-open-data-124356672deb

Cruz, M., & Lazarow, A. (2021, June 23). Innovation in government: Brazil. *McKinsey & Company*. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/innovation-in-government-brazil

Cyganiak, R., & Maali, F. (2015). Use cases and requirements for the data catalog vocabulary. *World Wide Web Consortium (W3C)*.

Darpan Dashboard. (n.d.). National Informatics Centre. Government of India. https://meity.dashboard.nic.in/login.aspx

Data Portal Asia. (2021). https://dataportal.asia/home

Davies, T. (2012). 5-Stars of Open Data Engagement. *Tim's Blog*.

Deloitte & Ministry of Housing and Urban Affairs. (2020). *Technology and Data Governance in Cities Indian Smart Cities at the Forefront of the Fight Against COVID-19*. World Economic Forum. https://www3.weforum.org/docs/WEF_Technology_and_Data_Governance_in_Smart_Cities_India_2020.pdf

Department of Science & Technology, Government of India. (2021). Draft National Geospatial Policy 2021. https://dst.gov.in/sites/default/files/Draft%20NGP%2C%202021.pdf

Drafting Team Metadata & European Commission Joint Research Centre. (2007). INSPIRE

Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 https://inspire.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf

Etalab. (n.d.). *Demander à data.gouv.fr de moissonner votre site.* data.gouv.fr. https://doc.data.gouv.fr/jeux-de-donnees/demander-a-datagouvfr-de-moisonner-votre-site/

Government of Brazil. (n.d.). *VLibras* https://www.vlibras.gov.br/

Government of France. (2021). *Code of relations between the public and the administration.* https://www.legifrance.gouv.fr/codes/id/LEGITEXT000031366350/

Government Open Data License. (2017). Ministry of Electronics and Information Technology, Government of India. https://data.gov.in/government-open-data-license-india

Government Valls I. (2014). *Public data policy*. Gouvernement.Fr. https://www.gouvernement.fr/en/public-data-policy

Grzenda, M., & Legierski, J. (2021). Towards increased understanding of open data use for software development. Information Systems Frontiers, 23(2), 495-513.

India Urban Data Stack Exchange. (2021). Ministry of Housing And Urban Affairs. https://iudx.org.in/

Jasinskaja, N. E.(2019). DCAT-AP - a key for European data publishers to reach the data consumers and create a European open data ecosystem. European Commission. https://ec.europa.eu/isa2/dcat-ap-key-european-data-publishers-reach-data-consumers-and-create-european-open-data-ecosystem_en

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2011). A classification scheme for open government data: towards linking decentralised data. *International Journal of Web Engineering and Technology*, 6(3), 266-285.

Maali, F., Cyganiak, R., & Peristeras, V. (2010, August). Enabling interoperability of government data catalogues. In *International Conference on Electronic Government* (pp. 339-350). Springer, Berlin, Heidelberg.

Máchová, R., & Lněnička, M. (2017). Evaluating the quality of open data portals on the national level. *Journal of theoretical and applied electronic commerce research*, 12(1), 21-41.

Metadata for data.gov.uk datasets. (2017). Open Government Data platform United Kingdom. https://data.gov.uk/dataset/b5e4be7e-6c36-41b6-9d48-cae195c32e34/meta-data-for-data-gov-uk-datasets

Ministry of Statistics & Programme Implementation. (2012). Guidelines for Statistical Data Dissemination. https://mospi.gov.in/documents/213904/0/Guidelines+for+Statistical+Data+Dissemination+%28GSDD%29%2C+February+2019.pdf/

Ministry of Transparency and Comptroller General of Brazil. (2018). *Brazil National Action Plan 2018-2021*. Open Government Partnership. https://www.opengovpartnership.org/members/brazil/commitments/BR0101/

Mission DATA. (2020). Government of France. https://www.numerique.gouv.fr/uploads/20200827_Plaquette_Techgouv_DINUM.pdf

Oliveira, M. I. S., de Oliveira, H. R., Oliveira, L. A., & Lóscio, B. F. (2016, June). Open government data portals analysis: the Brazilian case. In Proceedings of the 17th international digital government research conference on digital government research (pp. 415-424)

Open Data Portal Watch. (n.d.). https://data.wu.ac.at/portalwatch/

Open Government Data Platform Estonia. Government of Estonia. https://opendata.riik.ee/en/guidelines/

Open Government Data Platform France. (n.d.). Government of France. *Ask data.gouv.fr to harvest your site*. https://doc.data.gouv.fr/jeux-de-donnees/demander-a-datagouvfr-de-moisonner-votre-site/

Open Government Data Platform France. (n.d.). *Matomo*. Government of France. https://stats.data.gouv.fr/

Open Government Data Platform India. (2016). Release Of Compendium & Round Table Discussion To Reach 50,000 Datasets. https://event.data.gov.in/event/release-of-compendium-round-table-discussion-to-reach-50000-datasets/

Open Government Data Platform India. (2017). *#OpenGovernmentDataHack RoundTable Report*. National Informatics Centre. https://event.data.gov.in/wp-content/uploads/2017/07/OGD-RoundTable-12_Jul_17-Report.pdf

Open Government Partnership Bulgaria. (2011.) https://www.opengovpartnership.org/members/bulgaria/

Open Government Partnership Colombia. (2011). https://www.opengovpartnership.org/members/colombia/

Open Government Partnership Tunisia. (2018). National Action Plan 2018-2020. https://www.opengovpartnership.org/wp-content/uploads/2018/11/Tunisia_Action-Plan_2018-2020.pdf

Open Government Partnership. (2011). https://www.opengovpartnership.org/

OpenGovDataHack2020. (2020). Open Government Partnership Tunisia. http://www.ogptunisie.gov.tn/fr/index.php/lancement-du-hackathon-opengovdatahack2020-autour-de-la-reutilisation-des-donnees-publiques-ouvertes/

Organisation for Economic Co-operation and Development. (n.d.). *Open Government Data*. *OECD*. https://www.oecd.org/gov/digital-government/open-government-data.htm

Petychakis, M., Vasileiou, O., Georgis, C., Mouzakitis, S., & Psarras, J. (2014). A state-of-the-art analysis of the current public data landscape from a functional, semantic and technical perspective. *Journal of theoretical and applied electronic commerce research*, 9(2), 34-47.

Sangokoya, D., Clare, A., Verhulst, S., & Young, A. (2016, January). *Empowering Citizens to Make Data-Driven Decisions on Health Care*. GovLab. https://odimpact.org/files/case-study-uruguay.pdf

Satark Nagrik Sangathan. (2014). *Peoples' Assessment of the RTI Regime in India 2011–13*. RTI Assessment and Advocacy Group (RaaG) & Samya - Centre for Equity Studies. https://snsindia.org/rti-assessments/

Saxena, S., & Janssen, M. (2017). Examining open government data (OGD) usage in India through UTAUT framework. *Foresight*.

Sayogo, D. S., Pardo, T. A., & Cook, M. (2014, January). A framework for benchmarking open government data efforts. In *2014 47th Hawaii International Conference on System Sciences* (pp. 1896-1905). IEEE.

Sethi, T., & Prakash, M. (2018). Counting on Statistics: How can national statistical offices and donors increase use. *Williamsburg, VA: AidData at William & Mary*.

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government information quarterly*, 32(3), 308-315.

SmartCities Open Data Portal.(n.d.) Ministry of Housing And Urban Affairs & National Informatics Centre. https://smartcities.data.gov.in/

Solar, M., Concha, G., & Meijueiro, L. (2012, September). A model to assess open government data in public agencies. In *International Conference on Electronic Government* (pp. 210-221). Springer, Berlin, Heidelberg.

South-South Facility. (2012). Strengthening of Open Government in Tunisia. https://www.southsouthfacility.org/results/strengthening-open-government-tunisia

The Data Act. (2019). The Federal Government of United States. https://www.congress.gov/bill/115th-congress/house-bill/4174/text

The Federal Chancellery. (2019). *Second National Action Plan 2019–2021*. https://www.opengovpartnership.org/wp-content/uploads/2019/09/Germany_Action-Plan_2019-2021_EN.pdf

The Federal Ministry of the Interior. (2014). *The Federal Government's National Action Plan to implement the G8 Open Data Charter*. https://www.cio.bund.de/SharedDocs/Publikationen/DE/Aktuelles/20140918_nationeler_aktionsplan_open_data_englisch.pdf?__blob=publicationFile

The International Open Data Charter. (2013). https://opendatacharter.net/

Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives.

Umbrich, J., Neumaier, S., & Polleres, A. (2015, August). Quality assessment and evolution of open data portals. In 2015 3rd international conference on future internet of things and cloud (pp. 404-411). IEEE

UN Statistics. (2018). *Data Interoperability Guide. Annex A: A Roadmap to Interoperability*. https://unstats.un.org/wiki/display/InteropGuide/Annex+A%3A+A+Roadmap+to+Interoperability

van der Waal, S., Węcel, K., Ermilov, I., Janev, V., Milošević, U., & Wainwright, M. (2014). Lifting open data portals to the data web. *In Linked Open Data--Creating Knowledge Out of Interlinked Data* (pp. 175-195). Springer, Cham.

Verma, N., & Gupta, M. P. (2012). Open government data: more than eighty formats. In *9th International Conference on E-Governance (ICEG 2012)*, Cochin, Kerala, India.

Wright, G., Prakash, P., Abraham, S., & Shah, N. (n.d.). *Report on Open Government Data in India*. Centre for Internet and Society, Bangalore. https://cis-india.org/openness/publications/ogd-report/view

Yang, H. C., Lin, C. S., & Yu, P. H. (2015, September). Toward automatic assessment of the categorization structure of open data portals. In *International Conference on Multidisciplinary Social Networks Research* (pp. 372-380). Springer, Berlin, Heidelberg

Zuiderwijk, A., & Janssen, M. (2015, June). Participation and data quality in open data use: Open data infrastructures evaluated. In *Proceedings of the 15th European Conference on e-Government* (pp. 351-359)

Acknowledgements

We would like to thank Jagan Shah for reviewing and providing feedback on the paper. We are grateful to Hemant Adarkar, Samuel Neufeld, Sandeep Saxena, Santanu Ghosh, and Vaidehi Tandel for providing valuable insights. This paper has benefited from conversations with Gyanendra Badgaiyan and Niranjan Rajadhyaksha. We would like to thank Vikram Sinha and Sam Downes for reviewing the paper and providing editorial support. We are indebted to the Rockefeller Foundation, without whom this work would not have been possible.

About the Authors

Sridhar Ganapathy is a Senior Associate at IDFC Institute who works on data science. Prior to joining the Institute, he worked as an analyst with the learning transformation team at EZ Vidya Pvt. Ltd., supporting program design, impact assessment and behavioural analytics. Sridhar also worked at Tata Consultancy Services' Banking and Financial Services. His role at the organisation encompassed software design, database administration, and performance benchmarking. He also developed anti-money laundering solutions for financial institutions. Sridhar taught high school Mathematics and Science as a Teach for India Fellow.

Harsh Vardhan Pachisia is a Senior Associate and leads the data science unit at IDFC Institute. His research focuses on topics of urban development, data science and governance. Prior to joining IDFC Institute, he worked as an Economics Teaching Fellow at the NYU Stern School of Business. Harsh has also been a Research Assistant to several NYU professors analysing topics such as ‘Shadow Banking around the World’ and ‘Social Media Entrepreneurship’.

Rajeswari Parasa is a Senior Analyst at IDFC Institute who works on Data Science. Her interest lies in building open source tools for enhancing data-driven governance in Indian cities. Prior to joining IDFC Institute, Rajeswari worked as a Product Manager at a tech startup - Lakeer Foundation, where she worked on building open-source tools for visualising and interacting with city-data. She was an Urban Fellow at IIHS, Bangalore.

Isalyne Gennaro is a Contributor at Helios.do. She was an Associate at IDFC Institute. Her research focused on urban development and public policy in India. Before joining IDFC Institute, she studied Indian land utilisation at HCP Design Planning and Management in Ahmedabad and worked on urban development regulations.