

# India's open data initiative: Opportunity for states

**Point of View** 



April 2023 \_\_\_\_\_\_ kpm<u>g.com/in</u>

## Table of contents

Introduction	3
What is open data?	4
Need for open data adoption in the current scenario	5
Shift towards open data culture	6
Design and components of open data in India	8
Various levels of ODP maturity of Indian states	11
Our findings and insights	13
Advantages & innovative applications of OD for states	16
Conclusion	20

## **1. Introduction**



India is a recognised technology powerhouse—it has a rapidly growing market of digital consumers, with over 500 million internet users; the world's largest digital identity programme (Aadhaar), with over 1.2 billion people enrolled; and a thriving epayment ecosystem with an average of over 1 billion unified payment interface (UPI) transactions in a month. India's IT industry generates USD191 billion in revenues, with a strong talent base of 4.5 million engineers. The majority of the top 10 global systems integrators are based out of India and there are over 9,000 tech start-ups (1,600 of which are in the deep technology space) thriving in the country<sup>1</sup>.

Moreover, India's vast population is the biggest catalyst for tech adoption. Its immense appetite for digital services makes it incumbent on the tech ecosystem to develop solutions that are affordable, scalable and profitable. It will be exciting to see this rigour and energy for innovation and technology advancement. As this journey intensifies, there are several technologies that will be crucial for India to attain its trillion-dollar digital economy goal. Data is essential to the sector if India is to capitalise on new technologies. In the new digital world, the data owner will benefit greatly. India has already begun the process of digitising data and analysing it for data-driven decision-making. However, modern technologies are driven by high-quality data and are data hungry. To achieve our global aspirations, it is crucial to have a status check on our data system and clear roles of actors involved with proper checks and balances.

In 2022, the Government of India (GOI) introduced the Personal Data Protection Bill and Non-Personal Data Governance Framework. These progressive and detailed documents highlight the government's intent to enable and empower open data governance across the country.

The paper's objectives are to describe the Indian open data ecosystem, examine the adoption of the Open Data Movement and identify potential for the states to participate in open data.



### 2. What is open data?



Open data refers to the data that is freely available for anyone to access, use and share. This means that the data is not restricted by copyright, patents or other intellectual property rights. Open data is often provided by government agencies and other organisations as part of their commitment to transparency and accountability.

Open data can be used for a variety of purposes, including research, policy analysis and the development of new products and services. By making data available to a wide audience, open data enables greater collaboration and innovation and can help to promote economic growth and social progress.

Open data can be accessed in various formats, such as CSV files, XML documents and APIs. Some common types of open data include geographic data, demographic data, economic data and environmental data.

Open data initiatives often involve the creation of online portals

and other platforms that enable users to easily find, access and download data sets. These initiatives may also include the development of tools and resources that help users to understand and analyse the data as well as the establishment of standards and best practices for data management and sharing.

Overall, open data represents a powerful tool for fostering collaboration, innovation and transparency and can help to drive progress in a wide range of fields and sectors.

## 3. Need for open data adoption in current scenario

The adoption of open data is important in the current scenario for several reasons. First, open data can help to promote transparency and accountability in the government and other organisations. By making data available to the public, organisations can be held accountable for their actions and decisions. Second, open data can also facilitate innovation and economic growth. By making data available to researchers, entrepreneurs and other innovators, it can be used to develop new products, services and technologies that can drive economic growth. Finally, open data can help to improve the delivery of public services and support evidence-based decision-making. By making data available to those who deliver public services, it can be used to better understand the needs of the population and to design and implement more effective policies and programmes.

There are many examples of how open data can be used in conjunction with artificial intelligence (AI) and machine learning (ML) to solve problems and drive innovation:

- Predictive maintenance: Open data on equipment failure rates and maintenance histories can be used to train ML models to predict when the equipment is likely to fail, allowing for proactive maintenance to prevent disruptions.
- Traffic prediction and optimisation: Open data on traffic patterns and transportation networks can be used to train ML models to predict traffic flow and optimise routes for vehicles.
- 3. Disaster response: Open data on disaster locations, weather patterns and infrastructure can be used to train ML models to predict the likelihood and impact of disasters, allowing for more effective response and evacuation planning.
- Healthcare: Open data on patient outcomes and medical procedures can be used to train ML models to predict patient outcomes and identify best practices for treatment.
- 5. Agriculture: Open data on crop yields and weather patterns can be used to train ML models to predict crop yields and optimise irrigation and fertilisation.
- Environmental monitoring: Open data on environmental conditions, such as air and water quality, can be used to train ML models to predict and monitor environmental issues.

Five application domains for data sharing in the manufacturing sector, as shown in the table below, can be used to illustrate the potential of data sharing in creation of generative AI models to solve business problems.

#### Application domain value mechanism

S. No.	Application Domain	Value Mechanism
1	Enhance asset optimisation	by sharing and combining data of similar production equipment across companies to increase machine uptime and product quality.
2	Track products along the value chain	by sharing product location, time and quantity data to optimise and automate end-to-end processes.
3	Trace process conditions along the value chain	by sharing data on product and process conditions to create a continuous digital product record.
4	Exchange digital product characteristics	on product shape, geometry and composition to create a digital product twin and automate processes.
5	Verify provenance	by sharing data along the supply chain to ensure the origin of raw materials, the components and the products are as expected.

## 4. Shift towards open data culture



In India, government accountability has entered a new era because of digitisation efforts. With emerging international evidence of the vital role played by data as an enabler in driving public policy across its lifecycle, the central and state governments have paid significant attention to their data systems over the past two decades. Activities and results of key schemes and projects are being monitored intensely. Information and communication technology (ICT) systems, including DISHA, PRAYAS and the Output-Outcome Monitoring Framework (OOMF), are currently being used to enable intragovernment data exchange and integration.

On the policy, regulations and guidelines front, in 2022, the Indian government introduced the Non-Personal Data Governance Framework and the Digital Personal Data Protection Bill to enable data sharing, which have even given several recommendations for data-sharing purpose as listed below.

**1. Sovereign purpose:** Data may be requested for national security, law enforcement, legal or regulatory purposes.

- Core public interest purpose: Data may be requested for community uses/benefits or public goods, research and innovation, policy development, better delivery of public services, etc.
  - Specific data with commercial importance may be recognized as high-value datasets
  - Utilise data for research purposes.
  - Consider health sector as a pilot use case for Non-Personal Data Governance Framework.

**3. Economic purpose:** Data may be requested to encourage competition and provide a level-playing field or encourage innovation through start-up activities (economic welfare purpose) or for a fair monetary consideration as part of a well-regulated data market.

- Data request by start-ups/businesses
- Data request by data trustee/governments
- Setting up data and cloud innovation labs and research centres to develop, test and implement new digital solutions
- · Leverage data as training data for AI/ML systems

The Non-Personal Data Governance Framework also gives recommendations for the appropriate data-sharing mechanisms and checks and balances that should be put in place in order to ensure horizontal applicability of data-sharing principles to all non-personal data. It also recommends the establishment of a Non-Personal Data Authority and details out its roles and responsibilities.

The government has tried to incentivise data sharing via policy through the Non-Personal Data Governance Framework of 2022.

In order to ensure greater access to non-personal data in a systematised manner, creating a new category of business in India called 'data business' has been ideated in the Non-Personal Data Governance Framework. Data business is a horizontal classification and not an independent industry sector. Many existing businesses in various sectors collecting data beyond a threshold level will get categorised as a data business. The data business will have the following features:

- A new category/taxonomy of business called 'data business' that collects, processes, stores or otherwise manages data and meets certain threshold criteria.
- 2. Within India, data businesses will provide open access to meta-data and regulated access to the underlying data.
- 3. The compliance process will be lightweight and fully digital.

Several state governments are using dashboard-based analytical tools to display the work done by their various agencies (e.g., Pratibimba by Government of Karnataka), which are providing decision-makers with sophisticated information in the form of straightforward charts and figures. The need for even better data management was recognised by bureaucracy because of data digitisation, the introduction of new methodologies and the growing significance of data in public policy.

To achieve India's goal of Open Government and the Open Data Movement, the National Data Sharing and Accessibility Policy was unveiled in 2012. The same year, data.gov.in was developed to gather all relevant government data in one place for greater public use. The Indian government launched the Digital India programme in 2015 to ensure that people could access government services online. In 2015, the NITI Aayog's connected office—the Development Monitoring and Evaluation Office (DMEO)—came into existence. The office has been providing the government with rigorous, data-driven, citizen-centric and outcomes-driven programme management and policymaking since its establishment in 2015. GOI believes that access to meta-data will enable intragovernment data exchange and integration. The government strongly believes that meta-data sharing by data businesses, will spur innovation at an unprecedented scale in the country. One of the associated key objectives of data exchange is to promote and encourage the development of domestic industry and start-ups that can scale their data-based businesses.

For example, automobile companies may collect data about roads through various sensors. A start-up will know that this data is available based on the meta-data provided by automobile companies. The start-up can request for access for this data and can combine this data with public traffic data to create a solution for safest road routes for senior citizens. Similarly, a governmentfunded research lab may collect and publish data on air pollution across different locations in the city. The traffic department and a real-time navigation app may publish road traffic data. Looking at the pollution and traffic meta-data, a smart-city start-up may decide to create a solution for identifying safe and least polluted routes.

With increasing number of disjointed platforms and data systems, there arose a need to do a thorough analysis of the current state of each ministry's and department's data preparation to map out the future and identify ways to make improvements. Considering this, the Data Governance Quality Index (DGQI) project was started in 2021, with the goal of evaluating ministries'/departments' data readiness on a standardised framework to encourage healthy competition and cooperative peer learning from best practices.



## 5. Design and components of open data in India



The Open Government Data (OGD) Platform India, commonly identified with the domain data.gov.in, is a platform for the Indian government's open data programme. This portal promotes the use of open data by providing a single point of access to data from various government departments and agencies available to the public.

This data is typically made available through online portals or databases in formats like CSV, XLS, JSON, XML, RDF and others that are provided in an open format. Students, academics, businesspeople and members of civil society can utilise these accessible government datasets for research and development purposes that are both commercial and non-commercial and can be accessed by anyone with an internet connection.

Some key components of open data in India include:

- Government websites and online portals: These are the primary sources of open data in India. Government departments and agencies publish data on their websites, making it available to the public.
- 2. Datasets and databases: Open data in India is typically published in the form of datasets and databases. These can

be accessed and downloaded by anyone, allowing for further analysis and use.

- 3. Data standards and formats: To ensure that open data is easily accessible and usable, GOI has established standards and formats for the publication of data. This includes standards for the structure and content of datasets as well as formats for storing and sharing data.
- 4. Open data policies and initiatives: GOI has also implemented various policies and initiatives to promote the use of open data. These include initiatives to make data more accessible and user-friendly as well as policies to encourage the use of open data in various sectors.

Overall, the design and components of open data in India aim to make government data more accessible and usable, with the goal of promoting transparency and accountability in governance as well as driving innovation and economic growth.

### 5.1. Organisation and structure of open data portal (ODP)

The National Informatics Centre (NIC), Ministry of Electronics and Information Technology, GOI, is responsible for the design, development and hosting of the OGD Platform. The content published on data.gov.in website is owned by the relevant ministry, state, department or organisation and is available under the Government Open Data License, India.

Chief data officers (CDO) in states or ministries are empowered to nominate a few data contributors who would be responsible for contributing datasets and applications on the OGD Platform.



Source: KPMG in India Research

Though the portal is open for participation for all states, only a few states have made their representation on the platform and among these few only a handful of them are actively utilising the platform's features. There are also a few other states, which are not a part of the OGD Platform but have created their own variants of the portal for publishing their data with open licences.

	Updating ODP (Active)	Updating ODP (Less Active)	Updating ODP (Not Active)
Part of ODP	Karnataka Kerala Punjab Tamil Nadu	Madhya Pradesh Sikkim Kerala Meghalaya Himachal Pradesh Odisha Haryana Mizoram Chhattisgarh	Andhra Pradesh Assam Goa Gujarat Tripura Uttarakhand
Not part of ODP	Telangana (state level) Pune (city level)	-	Arunachal Pradesh Bihar Jharkhand Maharashtra Manipur Nagaland Rajasthan Uttar Pradesh West Bengal

Source: KPMG in India Research and India Open Government Data (OGD) platform

### 5.2. Observations on dataset quality

#### 5.2.1 Non-uniform data structures across states

NITI Aayog's DMEO closely monitors the data flowing from central ministries to maintain the quality of chosen national-level indicators to improve the ranking for Global Indices for Reforms and Growth of India. As a result, data flowing from these central agencies is more consistent and all state ODPs maintain the same indicator set.

The indicators, data formats and the level of granularity, on the other hand, are not consistent between states because the data coming from state agencies is collected in accordance with the needs of each individual state.

#### 5.2.2 Restriction on allowable data formats

The data platform offered by NIC allows for fixed format kinds as WMS, API, JSON, XML, XLS, ODS and CSV. Therefore, data owners are limited in their ability to publish data types in other machine-readable forms that are not supported by the platform.

Few states have captured geographic boundaries using shape files; some have mapped satellite imagery using Geographic Information System (GIS). While this kind of data is important for analysis, the lack of provisions on the ODP platform for data types is forcing data owners to create external links with restricted access.

#### 5.2.3 Published datasets have limited relevance for research

The idea behind OGD is that by making government data available to the public, entrepreneurs and innovators will have more opportunities to develop products that solve societal and organisational issues.

The data flowing into the portal must be clear, organised and, most crucially, of economic value if the goal is to be achieved. Companies will not be interested to utilise the disclosed data for

#### 5.2.4 States' alignment to OGD's larger agenda is limited

The potential of data and the influence it can have on governance for better service delivery and policymaking are only now being recognised by a small number of advanced states. Even though many states have developed digital systems, they are unable to extract value from the data they are gathering.

Additionally, the departments uploading data to the portal do not fully comprehend who and how the data will be used, which may be the cause of the platform's poor data quality. As on today, most of the state ODP data catalogues provide contact any kind of study unless the datasets have value for them financially.

This, therefore, makes the role of CDOs more crucial as they need to ensure that the ODP contains datasets with significant economic value and datasets that are easier to use for any kind of analysis.

details, address details and incentive disbursements at the scheme level.

OGD shouldn't be used in the same way as any other dashboard that combines data to display eye-catching graphics. The published datasets from different departments should be interoperable and the overlaying of different departments' data should yield less obvious insights.



## <u>6. Various levels of ODP maturity of Indian states</u>

An attempt has been made to compare different states on levels of maturity scale they fall under when it comes to publishing open data. As mentioned in the previous section, each state perceives utility of data differently and assigns prominence to ODP in that relative perceived priority order. In the process, states like Karnataka and Tamil Nadu appear to be above the others in a few dimensions.



DMEO, NITI Aayog, closely monitors data flowing from central ministries to maintain the quality of chosen national-level indicators, to improve the Global Indices for Reforms and Growth of India. As a result, data flowing from these central agencies is more consistent, and all state ODPs maintain the same indicator set.

	Geographic Granularity	Temporal Granularity	Recency of Datasets	Department Participation
Karnataka 🥊	High	High	Recent	Balanced
Tamil Nadu 🔰	High	Medium	Less recent	High
Kerala	Low	Medium	Less recent	Balanced
Sikkim	Medium	Medium	Old	Balanced
Madhya Pradesh	Medium	Medium	Old	High
Odisha 📌	Medium	Absent	Old	Balanced
Meghalaya 🗮	Low	Low	Old	High
Punjab 🥠	High	Low	Old	Balanced
Assam	Low	Absent	Absent	Low
Gujarat	Low	Absent	Absent	Low

Source: KPMG in India Research

The indicators, data formats and the level of granularity, on the other hand, are not consistent between states because the data coming from state agencies is collected in accordance with the needs of each individual state.

To further analyse factors that are contributing to a state's ODP maturity levels, these four qualifiers can be considered— geographic granularity, temporal granularity, recency of datasets and department participation.

**1. Geographic granularity:** This means that at what level of administrative unit is the dataset being made available. Datasets available at the smallest measurable administrative unit rather than available at an aggregate level will mean high granularity.

It is observed that from among the states that were compared, states such as Karnataka, Tamil Nadu and Punjab, were observed to define datasets with high granularity. These states are providing information below state aggregate data and are segregating datasets at district/regional levels.

2. Temporal granularity: This refers to how regularly is the dataset captured or published, such as daily, weekly, monthly, annual, etc. Datasets that are more frequently captured will mean high temporal granularity compared to the ones captured with annual or decadal frequencies.

It is observed that datasets in Karnataka ODP alone have datasets with high temporal frequency. Though the state has the majority of datasets published annually, recently the datasets are being captured and published on a monthly basis. There are also several indicators that are real time in the form of GIS representation, while other states are mostly publishing datasets at annual frequencies. **3.** Recency of datasets: This means how recent is the data that has been published. Datasets that are up to one year old may be considered as recent, datasets that are up to three-year-old may be considered less recent and datasets that are older than three years may be categorised as old.

Again, Karnataka ODP is the only state that has published more recent datasets that are as recent as few months. Tamil Nadu and Kerala have tried to publish datasets that are up to two years old and the rest of the states have datasets that are beyond even four years old.

**4. Department participation:** This refers to the number of participating state departments that are supplying datasets that are to be published on ODP. States with a large part of datasets coming from state-controlled departments are categorised to have high participation. While the states with equal contribution from state-controlled departments and Public Sector Undertakings (PSUs) towards supply of datasets to the ODP are categorised as balanced participation. States that have no state-controlled department participation but only PSUs contributing to the ODP are categorised as low participation.

Many states have balanced participation with PSUs and state departments contributing together for ODP. Tamil Nadu, Madhya Pradesh and Meghalaya have high state department participation.



### 7.1 Geographical granularity

In recent years, the openness of public datasets has gained a lot of traction and the most important parameter that has been considered by research institutes, private organisations and civil societies is the geographic granularity of the data and at what specific geographical level the data has been collected—at state, district, municipalities, ward, tehsil or a particular point level. But why is there so much focus on geographical granularity? The reason is obvious: to make decisions based out of data and that too for specific points of interest. Initiatives such as the India Election Dataset, Data in Climate Resilient Agriculture, India Driving Dataset are already live and have published such open data which has primarily focused on geographic locations. The data is intended for the citizens, policymakers, administrators, researchers and private organisations, to come up with advanced-level policies, projects, innovations and ideas that can be implemented for better public delivery system.

Similarly, on data.gov.in portal, some state governments have published certain datasets which are showcasing their granularity extent at state-, district-, tehsil-, village- or point-level data.

For analysis, we have categorised the granularity of the data on the following scale:

Granularity of Data	Scale
State portal having no data	Not applicable
State portal having state-level data	Low granularity
State portal having both district- and state-level data	Medium granularity
State portal having mandal-, block-, village- or geo-tagged data	High granularity

On the basis of this scaling, we have gathered the following insights.



### Geographical Granularity of the Dataset

Source: KPMG in India Research

#### **High granularity**

- 1. The state of Karnataka has published nine datasets of high geographical granularity. Of these, four are point-level location data, one is of block level and the rest four are at the mandal level. The state of Meghalaya has published one dataset of high geographical granularity provided at village level.
- 2. Odisha has published two datasets of high geographical granularity—one at the block and another at the mandal level.
- 3. The state of Sikkim has published two datasets of high geographical granularity at the block level.
- 4. Tamil Nadu has published four datasets of high geographical granularity—one is at the village level and the other three are at the block level.

#### Medium granularity

Karnataka (42 datasets), Punjab (51 datasets), Tamil Nadu (32 datasets), Sikkim (5 datasets), Madhya Pradesh (4 datasets) and Kerala (3 datasets) have medium granularity datasets both at the state and district levels. (Note: Meghalaya and Odisha do not have medium granularity datasets published on their ODPs.)

#### Low granularity

Karnataka (56 datasets), Tamil Nadu (34 datasets), Punjab (18 datasets), Sikkim (14 datasets), Kerala (7 datasets), Meghalaya (5 datasets and both Odisha and Madhya Pradesh (4 datasets) have low granularity datasets at the state level.

(Note: Gujarat and Assam do not have any dataset at any granularity.)

### 7.2 Temporal (time series) granularity

In order to analyse the effectiveness of the data parameters, it is important to understand at what time frame the data is being collected so that policies and implementation can be made according to it. For example, financial, agriculture, education, power and transportation sectors require monthly and daily time-frame datasets to analyse the pattern of the actual scenario and how the state can intervene to improve the trend and lay out better policy formulation.

In our study, we have analysed how frequently states have collected the datasets for different sectors and what their recency in comparison to the year 2022 are.



We have gathered the following insights.

Source: KPMG in India Research

1. Karnataka is the only state that has one real-time (COVID-19 monitoring), monthly and annual, monthly and biannual dataset on its portal.

2. In terms of annual data, Punjab is on top (datasets), followed by Karnataka (55 datasets), Sikkim (5 datasets) and Kerala (2 datasets).

3. Punjab (9 datasets) and Meghalaya (1 dataset) are the only states which have five years of dataset on their open data platform.



- Karnataka (71 datasets) and Tamil Nadu (3 datasets) are the only two states which have datasets that are less than one year old. Also, Karnataka (23 datasets) and Tamil Nadu (2 datasets) have datasets which are one year old. We, therefore, deduce that these two states are updating their data frequently.
- 2. Tamil Nadu (8 datasets) and both Karnataka and Kerala (4 datasets each) have datasets that have been updated two years ago.
- 3. Punjab (54 datasets) is the only state which has the maximum number of datasets that were updated three years ago followed by Kerala (2 datasets).
- 4. Only two states—Punjab (3 datasets) and Odisha (1 dataset)—have four-year-old datasets.
- 5. Sikkim (13 datasets) has datasets which were updated more than five years ago, followed by Punjab (12 datasets, dating back to the 1980s), Tamil Nadu (7 datasets), Madhya Pradesh (6 datasets) and Meghalaya (3 datasets).

### 8. Advantages and innovative application of Open Data for states



Given that the world has become more and more data-driven, open data is crucial. But the concept of data-driven business and governance will not be realised if there are limitations on the access and use of data.

The advantages of making data publicly available and simple to use can be substantial. There are several potential benefits of open data for governments, which are listed below:

- Improved transparency and accountability: Open data can help governments be more transparent by making it easier for citizens to see how their tax is being used and what their government is doing. This can help build trust between governments and the people they serve.
- 2. Increased efficiency and cost savings: Open data can help governments identify inefficiencies and areas where they can streamline processes, potentially leading to cost savings. For example, open data could be used to identify areas where there are duplicative services or opportunities to automate processes and reduce the need for manual labour.

- Economic development: Open data can be used to drive innovation and create new businesses and job opportunities. For example, open data on transportation patterns could be used to develop new transportation services or open data on health outcomes could be used to develop new healthcare technologies.
- 4. Improved decision-making: Open data can help governments make more informed decisions by providing them with access to a wider range of data and perspectives. For example, open data could be used to identify trends or patterns that might not be immediately apparent from a narrow set of data.
- 5. Enhanced public services: Open data can be used to improve the delivery of public services. For example, open data on crime patterns could be used to help allocate police resources more effectively or open data on health outcomes could be used to improve the delivery of healthcare services.

### 8.1 Testimonies of states that are benefiting from open data

Case 1: Advancing the state-of-AI research in road safety to reduce accidents and fatalities

Issue addressed: The increasing rate of road accidents and fatalities in the country

**Gap:** While several datasets for autonomous navigation have become available in recent years, they have tended to focus on structured driving environments. Driving conditions in India are quite diverse and traffic behaviour is highly unstructured compared with the rest of the world.

**Dataset collected:** Indian Driving Dataset (IDD) is a novel dataset for road scene understanding in unstructured environments. It consists of 10,000 images, finely annotated with 34 classes collected from 182 drive sequences on Indian roads. The dataset consists of images obtained from a front-facing camera attached to a car driven in Hyderabad and Bengaluru.

**Benefits to the state:** The images published in IDD are being utilised by start-ups to build AI models that can identify potholes in the road. This dataset is also providing opportunities for the global research community to investigate emerging AI concepts and benchmark their solutions.

Reference: Mobility dataset created by INAI, research center in association with IIITH

#### Case 2: Data in climate resilient agriculture

Issue addressed: Uncertainty in crop output and food security

**Gap:** There exists disparate databases on rainfall, cropping patterns, soil health and mandi pricing but none of these can be combined to advise farmers on the essential steps to take for a bountiful harvest and the best market price.

**Dataset collected**: Rastar and vector shape files were produced down to the mandal level for the whole state of Telangana. At the mandal level of detail, many layers of socio-economic, environmental and infrastructural data were also gathered throughout the state.

**Benefits to the state:** More insightful information was obtained by superimposing shape files with datapoints like soil moisture, weather information and land surface temperature. This information was used to advise farmers on the type of soil nutrition they should maintain, the type of irrigation practices they should adopt, the closest marketplace to which they should take their harvest and other things.

Reference: Data in Climate Resilient Agriculture dataset project supported by UNDP

Case 3: Open Energy project encouraging energy efficient investments

Issue addressed: Low utilisation and electricity wastages

**Gap:** There is no transparency of data on electricity consumed by each consumer. Hence electricity generation is more dependent on heuristics rather than on actual demand. Energy wastages that went unchecked and strained resources and the environment were also caused by the absence of patterns for energy consumption.

**Dataset collected:** The Open Energy project is a platform where 24 public institutions publish their electricity consumption (in kWh) monthly to make these data more transparent and enable users to monitor patterns. Public institutions publish the metrics of their electricity consumption on the Open Energy platform. Visualisations of these metrics, including numbers and graphs, are open for anyone to access.

**Benefits to the State:** Through the platform, governments, citizens and businesses have access to data that can help improve performance in terms of energy efficiency, resources and environmental protection. This facilitates the exchange of best practices, encourages energy efficiency investments, and supports responsible electricity consumption across society.

### 8.2 Underlying issues to address

Looking at the mechanisms behind the data-sharing areas, the following barriers to data-sharing collaborations can be observed:

#### 1. Trust related barriers:

- Fear of unintentionally giving away valuable or sensitive data about the business
- Fear of losing negotiation power or a competitive advantage
- Lack of visibility into data usage and analysis once shared

Different levels of maturity and understanding regarding the usage of open data and its potential advantages exist among data owners who oversee data collection, data cleansing and data publication. Data owners can be divided into three categories based on their level of maturity: beginner, intermediate and advanced.

The data owners must be onboarded along with CDOs (who are essential to publishing data onto the ODP) if we are to accomplish the desired interoperability of data between ministries/departments and extract creative solutions from the government datasets.

India's current stance on open data is that open data will inspire solutions that will address socio-economic issues. According to scholars, particular research calls for data in particular formats. If data is not available in such tailored forms, it cannot be used to test research problem hypotheses.

Therefore, it is necessary to improve the standards and data quality of Indian open data, which necessitates a shift in strategy. The bull needs to be practically tied before the cart for the wagon to continue moving. Therefore, prioritising research issues that were previously disregarded is required before the construction of datasets to produce usable datasets.

To facilitate the interoperability of datasets, there needs to be a facilitating institution that can gather problem statements, which

#### 2. Technical barriers:

- Risk of data breaches and losses
- Accessibility and interoperability issues that arise from combining data
- Different digital maturity levels among participants in the same solution
- Costs of switching technologies (or fear of technological lock-in)

are of interest to either the government, the corporate or the public. The institution can pass on the problem statements to research institutes, start-ups and corporates and seek proper data formats, which can allow them to conduct research on the topic. The data formats thus sorted may be passed on to the relevant departments who can mobilise their staff to collect data in the required formats before publishing on ODP.

To address these issues, successful collaborations use the following:

- 1. A clear value proposition and rationale for data sharing
- 2. Mutually beneficial agreements
- 3. Secure technologies and common standards



### 8.3 Frameworks available to start with

To help states understand the tools necessary to improve the quality of their data, several open data benchmarks are available, including the World Bank's Open Data Readiness Assessment (ODRA), the World Wide Web Foundation's Open Data Barometer (ODB) and the Open Knowledge Foundation's Open Data Index (ODI).

#### 1. World Bank's Open Data Toolkit

An ODRA methodological tool was created by the World Bank's OGD Working Group. It can be used to conduct an action-oriented assessment of whether a government or specific agency is prepared to evaluate, design and carry out an open data effort.

#### 2. World Wide Web Foundation, ODB

The objective of the ODB is to determine the genuine prevalence and impact of open data efforts globally. By combining contextual information, technical evaluations and secondary indicators, it uses an in-depth technique to analyse global trends and provide comparative data on governments and areas.

The barometer ranks governments on:

- Readiness for open data initiatives
- Implementation of open data programmes
- Impact that open data is having on business, politics and civil society

#### 3. Open Knowledge Foundation's ODI

The Open Knowledge Network manages the Global Open Data Index (GODI), which serves as an annual standard for the release of OGD. According to the Open Definition, the crowdsourced poll assesses how openly accessible government data is.

The GODI provides significant insights to government data publishers to determine where they have data gaps by having a tool that is controlled by civil society. Additionally, it demonstrates how to make data more useful and ultimately more significant. GODI, thus, offers significant feedback that governments ordinarily lack.

### 8.4 Measures for the states to seek benefits of open data

The four key enablers of successful data sharing and data integration that can help address the challenges faced by open data stakeholders are:

- 1. Selecting technologies
- 2. Using common standards
- 3. Building trust
- 4. Having legal and regulatory certainty

Hence, if Indian states want to reap the benefits of open data, they need to focus on the following:

#### 8.4.1 Focus on processes and structure workflow

1. Develop a training programme for department officials on the usage and advantages of open data.

2. Establish CDO structures and give them the authority to separate confidential data from data for public use.

#### 8.4.2 Focus on data standardisation

- 1. Introduce a rating mechanism for data catalogues to separate high-value datasets.
- 2. Simplify the procedure for pulling data straight from existing departmental data portals into the state ODP utilising API.
- 3. Establish uniform administrative codes for the purpose of

data gathering. For instance, the transportation department gathers data at the Regional Transport Office (RTO) level, the revenue department gathers data at the mandal/village level and the health and welfare department gathers data at the Primary Health Center (PHC)/ Community Health Center (CHC) level. Data overlaying and subsequently data interoperability can be accomplished if there is a single standard and uniform terminology for administrative units.

#### 8.4.3 Focus to improve overall data quality

- Create interoperability facilitation institution that can coordinate and guide departments with relevant data structures and format types.
- 2. Create problem-linked data catalogues with the aid of an organisation that facilitates interoperability.
- 3. Establish an ethical committee and develop data use standards to protect data from cybercrime.
- 4. Evolve a system of continuous monitoring to keep data updated and renewed.

### 9. Conclusion



In conclusion, the use of open data in India has the potential to bring numerous benefits to the country. By making government data freely available to the public, open data can promote transparency and accountability in governance and can also drive innovation and economic growth. GOI has made efforts to promote the use of open data, including by making data from various departments and agencies available to the public and implementing policies and initiatives to encourage the use of open data. Government policies and guidelines, such as the Digital Personal Data Protection Bill and Non-Personal Data Governance Framework, will be the enablers for open data in India. Similarly, state governments also need to come up with their respective guidelines for open data.

However, there are also challenges to the use of open data in India, such as the need to ensure the quality and accuracy of the data and to provide the necessary infrastructure and support to allow individuals and organisations to effectively use open data. To mitigate these challenges, Indian states must follow these eight factors of best practices of state ODP as listed below:

- 1. Data classification
- 2. Data security and privacy
- 3. Governance
- 4. Data interoperability
- 5. Data accessibility
- 6. Data lifecycle management
- 7. Data inventory/storage
- 8. Data quality

Overall, the use of open data in India presents both opportunities and challenges and will require continued efforts from all stakeholders to fully realise its potential benefits. The onus is now on the states to make India's open data initiative a success. Individual state's data polices on the above 8 factors will help in making this a success.

# Acknowledgements

We are extremely grateful to senior leaders from the industry, subject matter experts and KPMG in India team members for extending their knowledge and insights to develop this report.

## **Authors**

- Satya Sinha
- Harsha Vardhan Ampili
- Shubham Sharma
- Danish Hakim
- Amogh Bhatnagar

## **Design and Compliance**

- Shreya Chakraborti
- Saloni Poddar
- Sameer Hattangadi

## **KPMG in India contacts:**

#### Brijendra Kumar

Partner, Digital Government Advisory, Government & Public Services T: +91 9560075707 E: brijendrakumar@kpmg.com

Chandan Singh Director, Digital Government Advisory, Government & Public Services T: +91 9899528793 E: <u>chandansingh3@kpmg.com</u>

#### Vijay Nekkalapudi

Director, Economic Development Advisory, Government & Public Services T: +91 8121224567 E: vnekkalapudi@kpmg.com

Follow us on:

kpmg.com/in/socialmedia



**30** years and beyond

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act on such information without appropriate professional advice after a thorough examination of the particular situation.

KPMG Assurance and Consulting Services LLP, Lodha Excelus, Apollo Mills Compound, NM Joshi Marg, Mahalaxmi, Mumbai - 400 011 Phone: +91 22 3989 6000, Fax: +91 22 3983 6000.

© 2023 KPMG Assurance and Consulting Services LLP, an Indian Limited Liability Partnership and a member firm of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved.

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.

This document is meant for e-communications only.