Unlocking Potential of India's Open Data

(A NASSCOM initiative to drive innovation through
greater adoption of open government data platform)

In collaboration with
Fractal, Microsoft, Infosys, IDFC Institute, TCS & Amazon

Foreword

With the pervasiveness of technology in our everyday lives and the unprecedented pace of digitalization, open data has become an asset with immense potential to catalyze research and innovation. The crux of any innovation is data, and as the Hon'ble Prime Minister, Shri Narendra Modi has succinctly stated, **data is the new gold.**

Data is a valuable economic and social resource offering enormous opportunities for citizens, businesses, and governments to make better-informed decisions and develop innovative products and services. Open data further promotes increased civil discourse, improved public welfare, and more efficient use of public resources. Furthermore, information becomes more valuable when shared and used appropriately, and in keeping with this philosophy, the Government of India launched the Open Government Data (OGD) Portal (Data.Gov.in) in compliance with the National Data Sharing & Accessibility Policy (NDSAP) in 2012. The aim of the OGD portal is to provide proactive access to Government owned shareable data along with its usage information in open/machine-readable format through a state-of-the-art platform.

However, to ensure quality datasets and information is usable by the ultimate consumer, there exists a need to understand the current demand and supply side challenges in the overall open data ecosystem.

In this context, it brings me great pleasure to see NASSCOM striving to bridge this gap between data providers and data users. NASSCOM, in collaboration with Industry partners, has published this report to identify the existing challenges in the open data ecosystem and has recommended ways and means to harness the potential of non-personal data.

The recommendations and action points in this report will benefit all stakeholders across the open data ecosystem and will create value and enable the implementation of research and innovation initiatives, including AI initiatives. Implementation of the relevant recommendations may add significant value to the existing efforts of transforming the open data ecosystem and ensuring that key stakeholders have access to quality datasets.

Mr. Abhishek Singh, IAS
MD and CEO, Digital India Corporation

Table of contents

01 Executive summary

02 Framing the problem

03 Methodology & approach

04 Insights & findings

05 Recommendations & roadmap

06 Appendix

01 Executive summary

Data is a strategic asset that has immense potential to drive innovation in today's tech-fuelled world. The unprecedented pace of digitisation and digital penetration in India has put it in a unique position to leverage the potential of this data to catalyse inclusive growth, research and transform public service delivery.

The government of India has millions of datasets. In accordance with the provisions of the National Data Accessibility & Use Policy 2012 (NDSAP), some of these datasets that are classified as open are currently published on the Open Government Data Portal (OGD).

However, these datasets and the platform currently suffer from a myriad of challenges ultimately impacting the usability of these datasets and the portal and keeping the data-driven research and innovation ecosystem from growing to its full potential in India.

Any open government data initiative must focus only on the most useful datasets, so as to not "boil the ocean". These highly valuable datasets are the focus of our report. For the purpose of this report, we define High Value Datasets (hereafter referred to as HVDs) are those that have potential for social impact. They are those datasets that are most useful to the users of the platform. As defined by the the Committee of Experts (Committee) on Non-Personal Data Governance Framework, an HVD is a dataset that is beneficial to the community at large and shared as a public good, subject to certain guidelines pertaining to the management of an HVD and data sharing.

This is a NASSCOM (National Association of Software and Service Companies) effort with industry participants that has been supported by NASSCOM and has consulted personnel from India's Ministry of Electronics & Information Technology (MeitY) for this project.

This effort involved extensive work to understand the gaps, issues and merits related to the platform and the data itself. From in depth interviews with various users, to a heuristic analysis of the platform and data audits to arrive at holistic recommendations that can help unlock value from India's data.

This report outlines the study and its conclusions.

Problem Framing

Data shared thus far on OGD is a subset of all the useful data that is available with the government of India. All data available to the Indian Government can be categorised as sensitive/non-sensitive, structured/unstructured, useful/not useful.

1.Drive adoption by sharply focusing on the top 10% datasets that have the most value. We will call these HVDs.

An HVD should:
i. Be useful for policy making and improving public service and citizen engagement
ii. Help create new and high-quality jobs
iii. Help create new businesses – start-ups and SMEs
iv. Help in research and education
v. Help in creating new innovations, newer value-added services / applications

Published HVDs

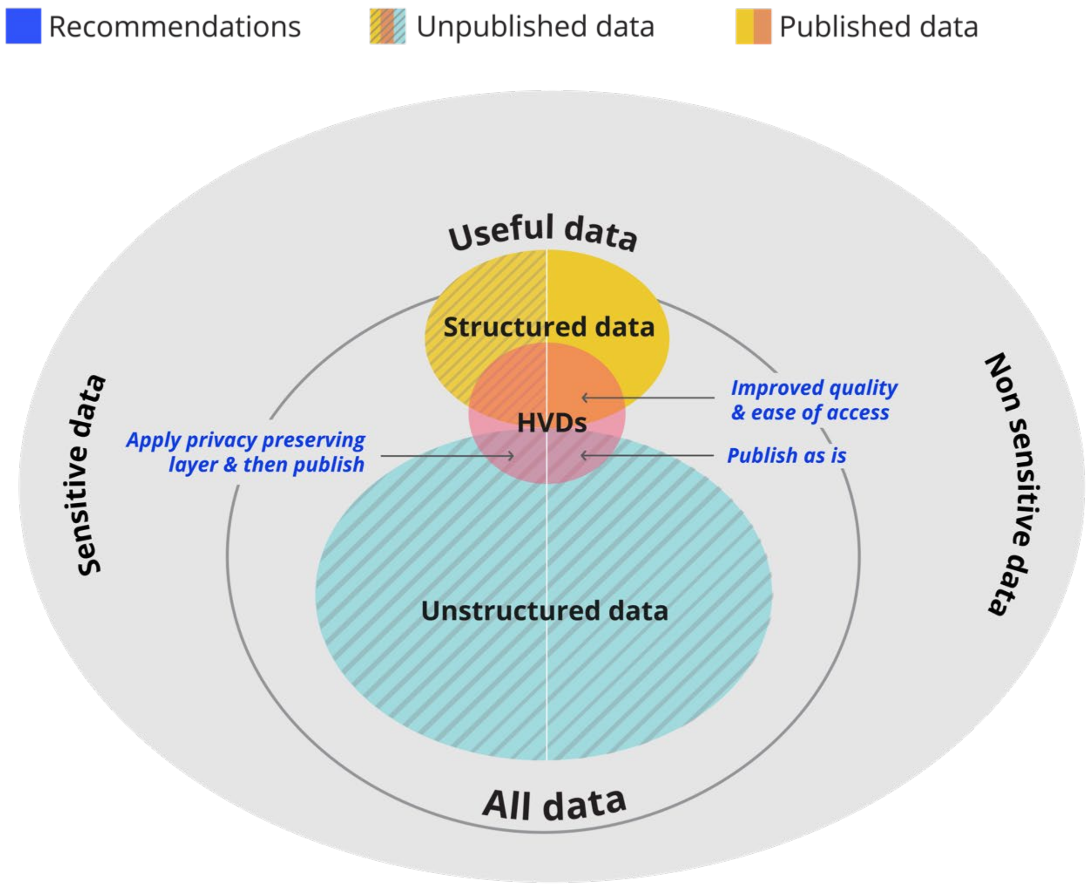
2. Of all the HVDs already published, drive ease of use and quality/reliability of the data. This will drive significant adoption.

Unpublished HVDs

3. Of the remaining high value datasets, publish as-is datasets that are previously unpublished because they were deemed machine not-readable. All datasets are machine-readable in the age of AI.

4. Of the remaining high value datasets, for those previously unpublished because they were deemed sensitive, apply a privacy preserving layer and then publish the same.

Figure below provides an indicative view of Government's data landscape and our overall recommended approach on unlocking useful data in each category:



Solution approach:

The taskforce constituted of industry leaders with an avid interest in minimizing roadblocks in the innovation ecosystem. The taskforce was active for a span of 6 months where various working groups were constituted to deep dive into specific components such as auditing the open data platform, developing a growth strategy, governance and expanding data coverage. Respective groups deep dived in the OGD ecosystem to understand the drivers of success of the OGD platform and the needs/perspectives of all the relevant stakeholders (including users, chief data officers (CDOs), NIC team, etc).

Recommendations

Our recommendations can be classified into two main categories
1) Increase HVDs on OGD and 2) Drive platform adoption

1. Increase HVDs on OGD

Discover high value dataset (HVDs): We recommend adopting the HVD framework shared in the report to identify, manage & monitor HVDs for driving engagement on OGD. The framework identifies HVDs based on an integrated scoring criterion, based on use case, usage and, market value parameters. We also recommend a classification criterion through the development of an AI/ML system, that will auto update the input parameters in the score. The framework also includes recommendations on regular maintenance of the HVD List and robust data governance that ensures high data quality along with periodic updates communicated to users.

Enriching existing HVDs: We recommend enriching the 19 existing HVDs on OGD platform and monitoring their quality regularly based on data quality parameters mentioned in this report. The data quality gaps in these 19 HVDs as shared by the OGD team spans across sectors of agriculture, census, shipping, healthcare, finance and telecom. These gaps pertain to completeness, consistency, timeliness and relevancy of datasets are deterrents to user adoption and due course correction is required.

Augment HVD list: It is also recommended that the HVD list be augmented, by including the 93 datasets that were identified from the extensive user research in combination with the HVD framework. CDOs in each ministry are recommended to find comparable datasets, through internal discovery using the HVD framework and enable access to them via the OGD platform.

Privacy preservation: Privacy preserving technologies should be made available to CDOs on the OGD platform, empower them to share data in a more secure fashion with due guidelines to mask any identifiers. We recommend in this report a privacy preserving framework based on mathematical measures that adopts techniques such as differential privacy and multi-party computation. This framework is to be applied on datasets to ensure privacy without any data distortion and minimum ambiguity. It is also recommended that the Privacy Preserving Framework be integrated with Cyber Security Framework to support the Data Exchange Ecosystem in managing privacy risks.

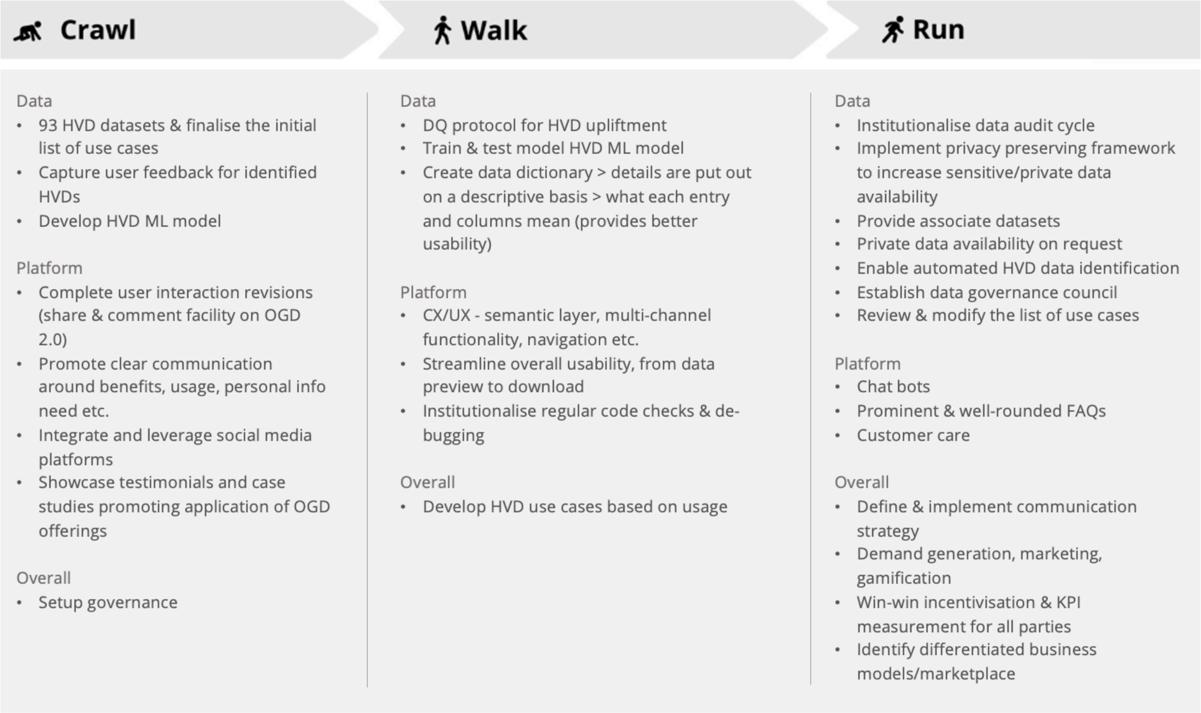
Model data sharing framework: Creating transparent, standardised, and risk-based data classification guides along with model data sharing framework is advised. Such an exercise can provide clarity on data classification and also facilitate CDOs to publish more datasets on the platform.

2. Drive adoption of platform

The Open Government Data platform has been operational for close to a decade and has served as the central destination to host a plethora of datasets by ministries, states and departments. This has been critical in enabling accountable governance as well as inform research and innovation in a variety of sectors. However, the taskforce has identified that there is still immense scope to enhance adoption of the platform especially by the innovation community.
Streamlining user navigation, reducing information overload, adding community engagement features and establishing feedback loops have been identified as immediate measures that can enhance overall user experience of the OGD platform. Integration of the OGD platform with social media websites, showcasing of testimonials and case studies, provisioning for assistance via chatbots, customer care support & prominent FAQs are key recommendations for a seamless user experience that will also build trust. Wireframes capturing these recommendations are included in the report.

Roadmap

All solutions, both incremental & strategic, are prioritised by data, platform & overall governance for relevant stakeholders into a roadmap.



Crawl

Data
- 93 HVD datasets & finalise the initial list of use cases
- Capture user feedback for identified HVDs
- Develop HVD ML model

Platform
- Complete user interaction revisions (share & comment facility on OGD 2.0)
- Promote clear communication around benefits, usage, personal info need etc.
- Integrate and leverage social media platforms
- Showcase testimonials and case studies promoting application of OGD offerings

Overall
- Setup governance

Walk

Data
- DQ protocol for HVD upliftment
- Train & test model HVD ML model
- Create data dictionary > details are put out on a descriptive basis > what each entry and columns mean (provides better usability)

Platform
- CX/UX - semantic layer, multi-channel functionality, navigation etc.
- Streamline overall usability, from data preview to download
- Institutionalise regular code checks & debugging

Overall
- Develop HVD use cases based on usage

Run

Data
- Institutionalise data audit cycle
- Implement privacy preserving framework to increase sensitive/private data availability
- Provide associate datasets
- Private data availability on request
- Enable automated HVD data identification
- Establish data governance council
- Review & modify the list of use cases

Platform
- Chat bots
- Prominent & well-rounded FAQs
- Customer care

Overall
- Define & implement communication strategy
- Demand generation, marketing, gamification
- Win-win incentivisation & KPI measurement for all parties
- Identify differentiated business models/marketplace

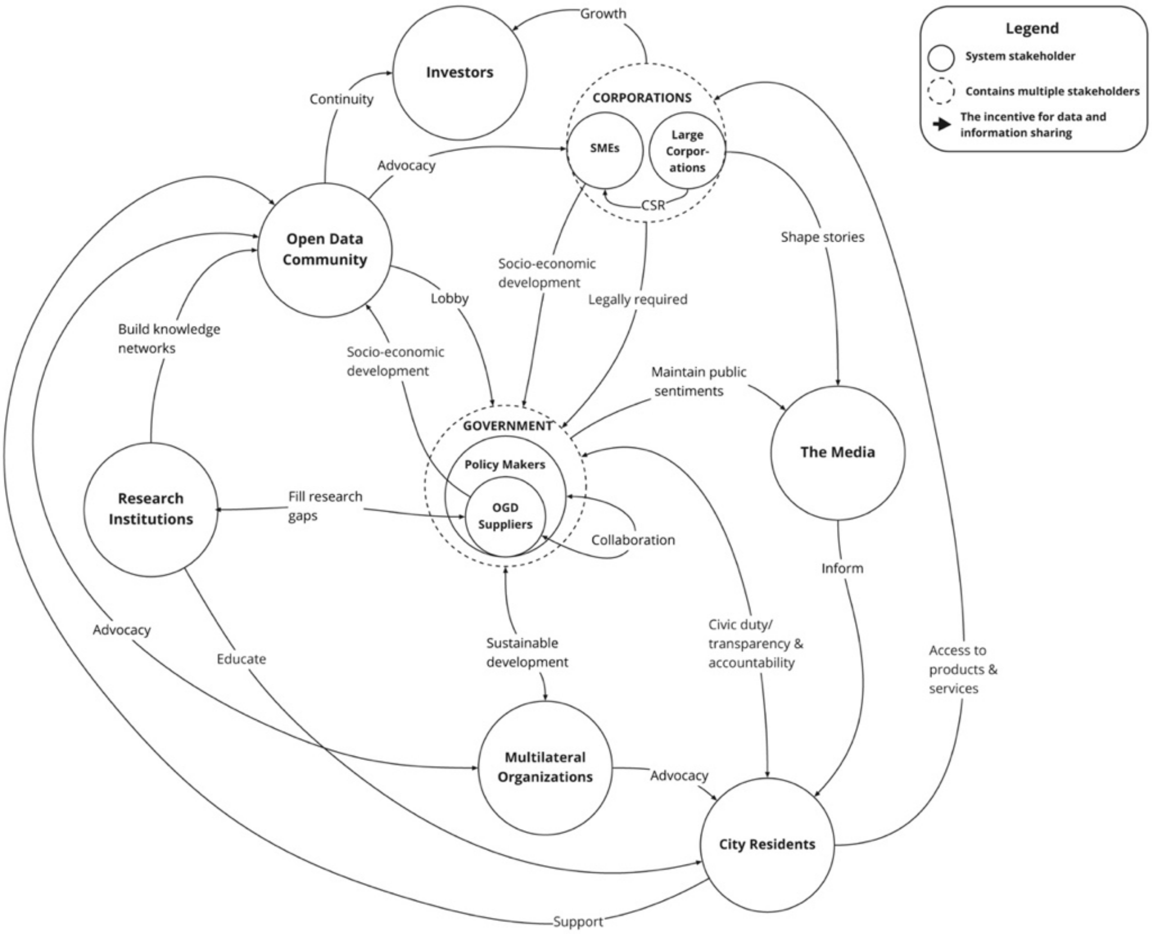
This report attempts to provide holistic recommendations to transform India’s data ecosystem.

02 Framing the problem

The following section gives a glimpse of all the working parts of the data ecosystem.

Stakeholder map for OGD

This map shows the various stakeholders who participate in the data sharing ecosystem, their motives and mandates. As is evident, a central unifying platform like OGD forms the nucleus of this network, providing critical avenues for data sharing.



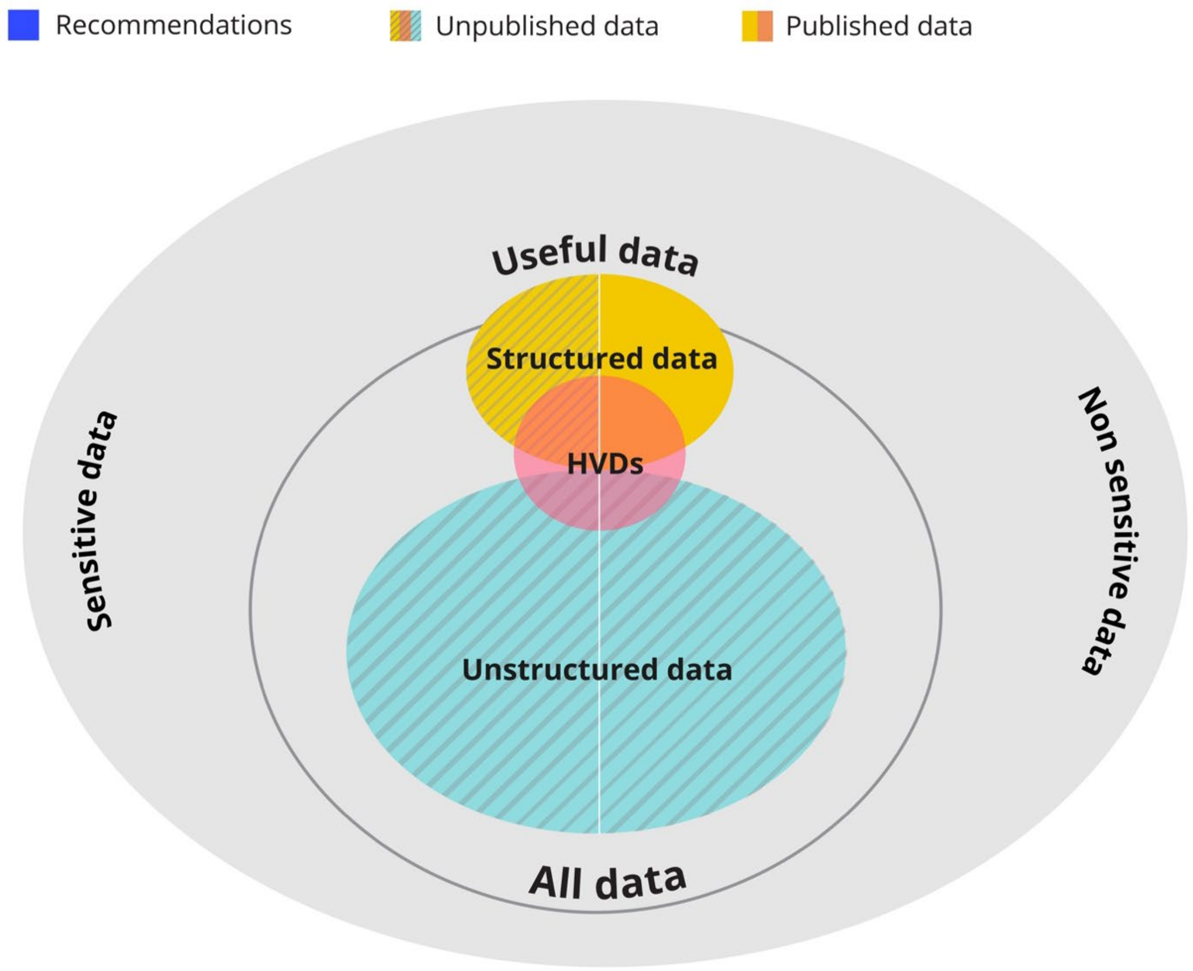
Chafetz, Hannah (2021) *Humanizing Data: A framework for Open Government Data decision making.*

Data ecosystem

All data available to the Indian Government can be divided into following sub-categories:

1. Usable data (subset of all data, public, potentially valuable)
2. Sensitive data (not made public)
    - Structured sensitive data (not public)
    - Un-structed sensitive data (not public)
3. Non-sensitive data
    - Structured non-sensitive data (public)
    - Un-structed non-sensitive data (non-public)

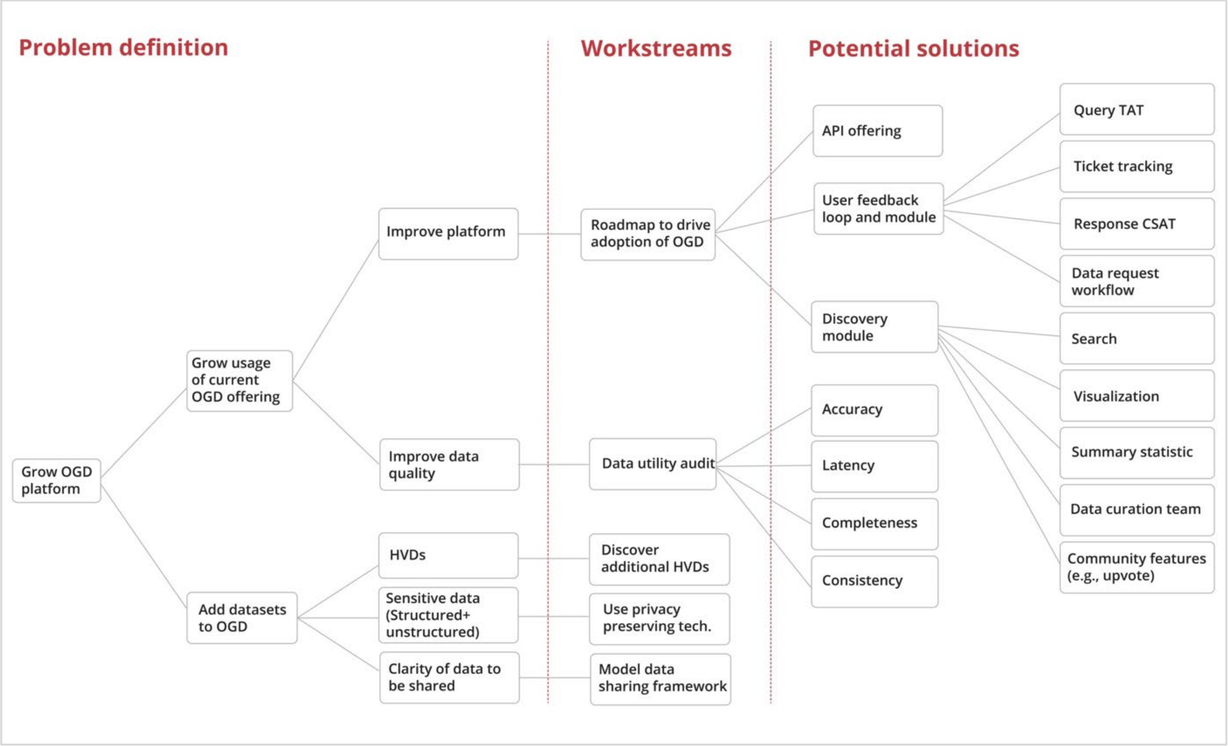
In our experience HVDs and high value use cases are typically 80% unstructured vs 20% from structured data.



Data shared in public domain is a subset of structured and unstructured data available

Problem unpacking

The strategy to drive growth of the OGD platform required a comprehensive needs assessment of the current data sharing landscape. To unpack the various layers involved, the following issue tree was mapped to break down the target into its constituent components.



Issue tree depicting the problem and its potential solutions

Takeaways

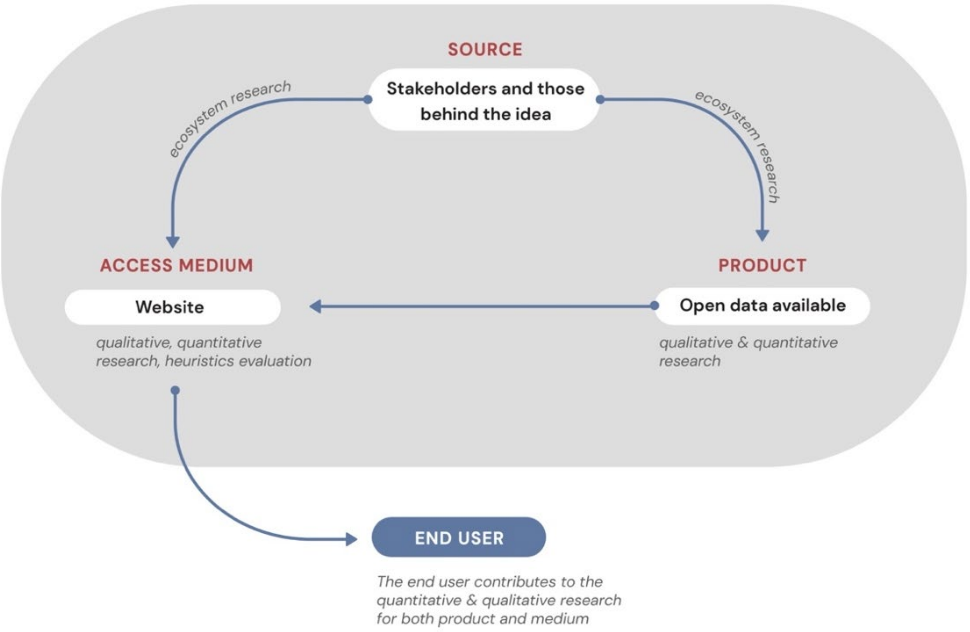
- Drive consumption of data.gov.in through right product management of the platform, categorization of the data available, and increasing granularity of data.
- Focus on top 10% datasets to improve adoption of public datasets by seeking feedback.
- Open-up a wider range of data with help of data anonymization tools.
- Provide access to all non-sensitive unstructured data.
- Identify high value datasets though a use case based approach by user research e.g. surveying relevant ministries/departments, industry bodies like NASSCOM, etc.

03 Methodology & approach

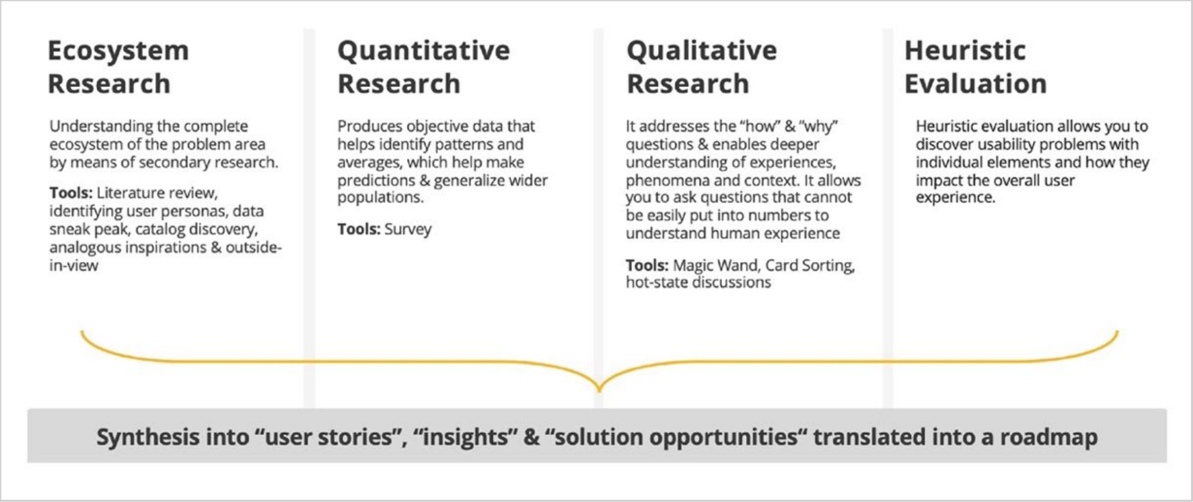
To understand the potential for growth of the OGD platform, we conducted both quantitative and qualitative research to understand the enablers and barriers to the uptake of the offerings. In addition, we conducted heuristic evaluation to identify the gaps on the website that hampered the access to the offerings (open data)

Research methodologies

We used a mix of 4 methodologies to understand the entire OGD ecosystem in depth. These methodologies have contributed to core insights, which in turn have helped shape a well-defined problem statement. We translated these findings into ‘user stories’, ‘insights’ & ‘solution opportunities and a roadmap.


SOURCE
Stakeholders and those behind the idea
ecosystem research
ecosystem research
ACCESS MEDIUM
Website
qualitative, quantitative research, heuristics evaluation
PRODUCT
Open data available
qualitative & quantitative research
END USER
The end user contributes to the quantitative & qualitative research for both product and medium

The diagram is a high-level representation of the OGD ecosystem, & the intersections at which we conducted enquiries



Ecosystem Research

Understanding the complete ecosystem of the problem area by means of secondary research.

Tools: Literature review, identifying user personas, data sneak peak, catalog discovery, analogous inspirations & outside-in-view

Quantitative Research

Produces objective data that helps identify patterns and averages, which help make predictions & generalize wider populations.

Tools: Survey

Qualitative Research

It addresses the "how" & "why" questions & enables deeper understanding of experiences, phenomena and context. It allows you to ask questions that cannot be easily put into numbers to understand human experience

Tools: Magic Wand, Card Sorting, hot-state discussions

Heuristic Evaluation

Heuristic evaluation allows you to discover usability problems with individual elements and how they impact the overall user experience.

Synthesis into "user stories", "insights" & "solution opportunities" translated into a roadmap

The 4 research methodologies we used

User archetypes

Every service, experience or product has different stakeholders and end users. Hence, we identified both direct and indirect users who are involved with this platform, to provide a well-rounded perspective to our approach.

A user archetype is a classic representation of a user, that outlines their usual behavior and characteristics. User archetypes form an important component of any user focused approach. We engaged with 51 users, a mix of data scientists, start-up founders, digital heads, agricultural scientists, policy makers, professors, advisors, and PhDs and split them into the following buckets:

Existing users
(n=16)

- Regular User
- Recent User
- Successful User

Non-users & prospective users
(n=31)

- Unaware of platform (interested)
- Aware of platform (not interested)
- Educational Institutions

Disengaged users
(n=4)

- Forgotten
- Uses other sources
- Dissatisfied experience

Data audit

We also conducted a detailed data audit for the 19 datasets on the OGD platform that were tagged as 'high value dataset's (HVD). This was done to understand the intrinsic aspects contributing to the supposed value of datasets. These datasets were further analyzed for inconsistencies in the provisioned data, availability of multi-format support, etc.

The following parameters were used to assess them:

- **Relevancy:** The data should meet the requirements of the intended use.
- **Completeness:** The data should not have missing values or miss data records.
- **Timeliness:** The data should be up to date.
- **Consistency:** The data be shared in a standard format to ensure that it can be cross referenced with the same results.

HVD evaluation

The overall objective of the high value datasets (HVD) framework is to help drive innovation and increase the usage and adoption of the Open Government Data platform.

Towards achieving this, we intend to deliver a HVD framework, which was designed using the approach mentioned below:
The parameters used to identify high value datasets did not include user demand or market value, and instead focused on data quality and other cosmetic parameters only. These parameters were solely reliant on the data providers and OGD team while not assessing the requirements of the demand for data.

Analyzed different HVDs across the globe to arrive at a framework, robust enough to compete with international best practices, the following were studied:

a. **European Union HVD framework**[1] - According to the directive, 'high-value' means data with the potential to generate significant socio-economic or environmental benefits and innovative services and shall benefit a high number of users and assist in generating revenues. It also provides six thematic

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELLAR%3Aa6ef4c41-97eb-11e9-9369-01aa75ed71a1

categories i.e., geospatial, earth observation and environment, meteorological, statistics, companies and company ownership and mobility.

b. **Canada open government framework[2]-** It has provided the common criteria to help identify high value datasets that can help the government in the following ways:
   - Helps identify social, environmental, and economic conditions
   - Helps promote better outcomes for public services
   - Encourages innovation and sustainable economic growth
   - Increases transparency, accountability, and the flow of information
   - Is in high demand by the community

c. **Australian framework[3]-** A user-centric approach, focused on understanding how businesses and not-for-profits are using public data., Stimulating use and re-use of public data to create social value, providing access to, and encouraging the use of public data and identifying and addressing barriers impeding the sharing of and access to data.

d. **Other Industry Frameworks[4]-** The objective was to define concrete high value datasets that fall under the thematic categories and focus on the 6 macro characteristics of HVDs that includes potential to generate social, economic, and environmental benefits, generate innovation/AI and improve/strengthen/ support public services and public authorities.

   Based on the analysis of these frameworks adopted by other countries, we understand that most of the frameworks focus on the HVD identification methodology and do not provide an exhaustive framework which covers management & monitoring of these high value datasets.

04 Insights & findings

Various methods were used to map the insights discovered during the taskforce's investigation. The core findings are illustrated in this segment.

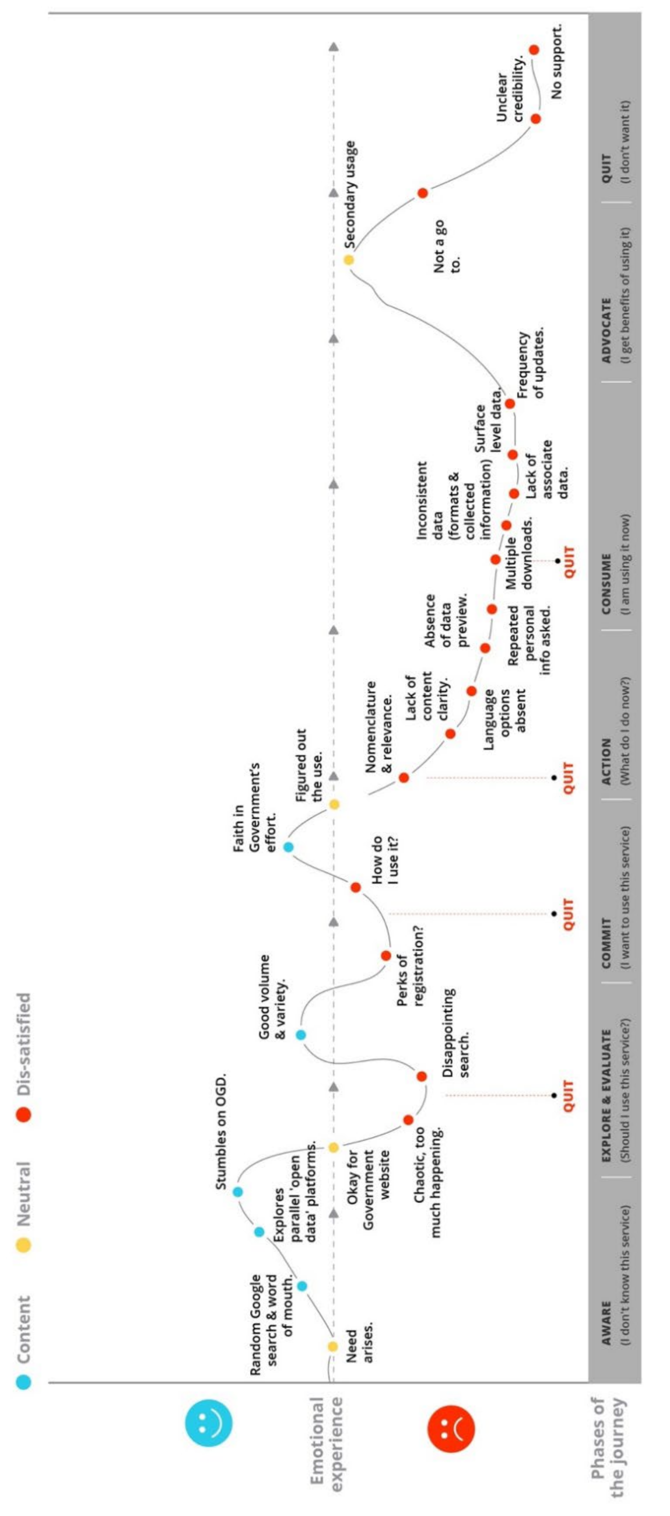
[2] https://opencanada.blob.core.windows.net/opengovprod/resources/2ac8c135-93a3-4fd7-88b6-ff870143276d/yopen-data-working-groupopeninfoaccessiblepdf-scanadian-open-government-working-group-high-value-.pdf?sr=b&sp=r&sig=nSZGxsy4EY97jEZZYkDH01m3CMiUNERSMMx/bAYtfjY%3D&sv=2015-07-08&se=2022-08-17T11%3A26%3A37Z

[3] https://ogpau.pmc.gov.au/national-action-plans/australias-first-open-government-national-action-plan-2016-18/21-release-high

[4] https://www.data.gv.at/wp-content/uploads/2020/02/Presentation-of-the-HVD-for-PSI-Study_MS-webinar.pdf

Journey map

From the conversations with various users, a journey-map was created to depict moments of truth within the ideal expected journey of a user on the OGD platform. The map represents the insights in the context of the user's journey through the OGD platform.

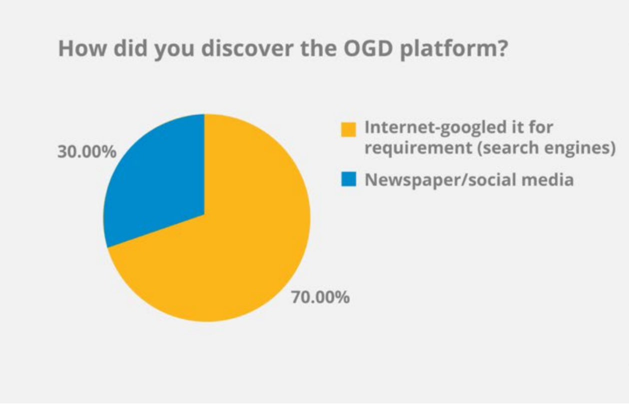

● Content
● Neutral
● Dis-satisfied
Emotional experience
Phases of the journey
Need arises.
Random Google search & word of mouth.
Explores parallel 'open data' platforms.
Stumbles on OGD.
Okay for Government website
Chaotic, too much happening.
Disappointing search.
Good volume & variety.
Perks of registration?
How do I use it?
Faith in Government's effort.
Figured out the use.
Nomenclature & relevance.
Lack of content clarity.
Language options absent
Absence of data preview.
Repeated personal info asked.
Multiple downloads.
Inconsistent data (formats & collected information)
Lack of associate data.
Surface level data.
Frequency of updates.
Secondary usage
Not a go to.
Unclear credibility.
No support.
QUIT
QUIT
QUIT
QUIT
AWARE (I don't know this service)
EXPLORE & EVALUATE (Should I use this service?)
COMMIT (I want to use this service)
ACTION (What do I do now?)
CONSUME (I am using it now)
ADVOCATE (I get benefits of using it)
QUIT (I don't want it)

Journey map illustrating end users emotional experience throughout the process, from awareness of OGD to end use of the datasets available.

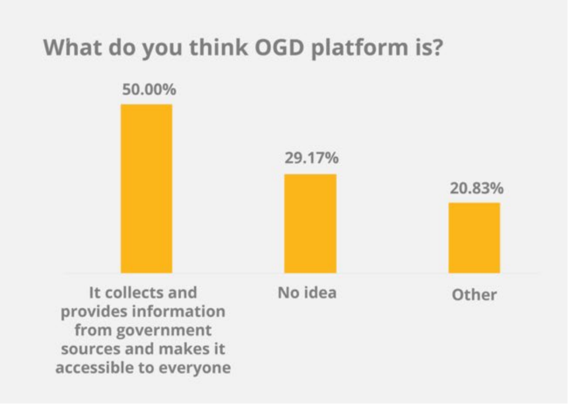
Awareness about OGD & it’s access

The following are curated list of findings and insights pertaining to the perception and awareness of open government data & the OGD platform. This is as discovered via interviews with various types of users.

- ‘Unstructured Google search’ and ‘word of mouth’ emerged as the main gateways to this platform’s discovery.
- Unclear value proposition of the OGD platform.
- ’Open’ in OGD has fluid connotations for users.
- Communication around OGD, that it is platform agnostic, accessible across devices.
- OGD is not the choice destination for any of their data needs, i.e they don’t come here directly, they usually stumble upon it.



The survey highlighted Google as the top access point along with social media and print.



50% of prospect and non-users don’t know what OGD is.

“Show value upfront.“

“A good secondary repository, a good to have platform for the users due to ineffective usage of the datasets.“

“I was googling and came across the hackathon, hence got to know about it.”

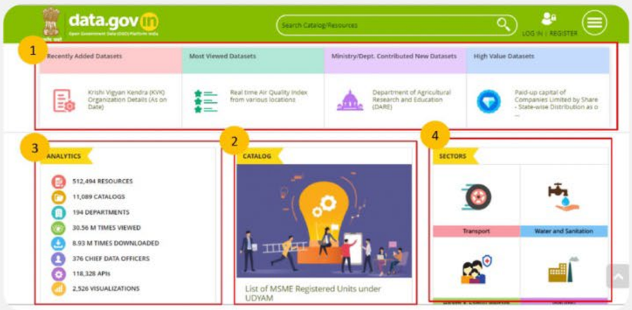
“The intent is good, but the platform must be focused, not arbitrary.”

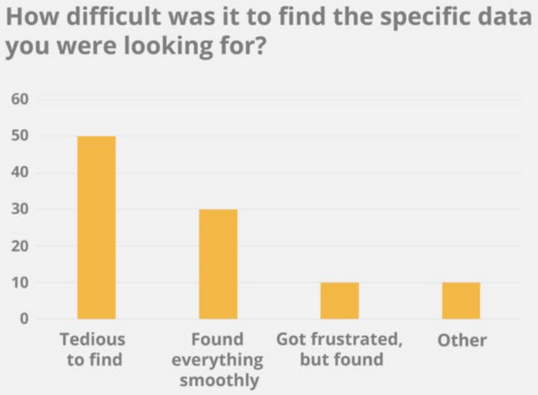
"How open this open data is going to be?" (In the industry like ours (micron) - it is hard to acquire the data itself)

The OGD platform

The following is a curated list of findings & insights pertaining to users' interaction and experience with the open government data website.

- OGD platform carries the baggage of the Government website experience.
- Lack of structure leads to navigational issues, search is not intuitive, purpose of registration is unclear.
- Frequent permissions & personal detail inputs to access data & APIs cause disengagement.
- Absence of language options.


data.gov.in
Open Government Data (OGD) Platform India
Search Catalog/Resources
LOG IN | REGISTER
1
Recently Added Datasets
Most Viewed Datasets
Ministry/Dept. Contributed New Datasets
High Value Datasets
Krishi Vigyan Kendra (KVK) Organization Details (As on Date)
Real time Air Quality Index from various locations
Department of Agricultural Research and Education (DARE)
Paid-up capital of Companies Limited by Share - State-wise Distribution as o ...
3
ANALYTICS
512,494 RESOURCES
11,089 CATALOGS
194 DEPARTMENTS
30.56 M TIMES VIEWED
8.93 M TIMES DOWNLOADED
376 CHIEF DATA OFFICERS
118,328 APIs
2,526 VISUALIZATIONS
2
CATALOG
List of MSME Registered Units under UDYAM
4
SECTORS
Transport
Water and Sanitation



1. Unclear nomenclature. 2. Prioritisation / description of the features are not self-explanatory.
3. Lacking clear 'call to action' overall.

Only 30% users were able to find what they were looking for, smoothly

"It is decent for a government website"

"Right now it looks like a data dump"

"How will a farmer use this, who will help?"

"Sometimes search works and sometimes it doesn't work. Sometimes it won't produce results that one might be looking for"

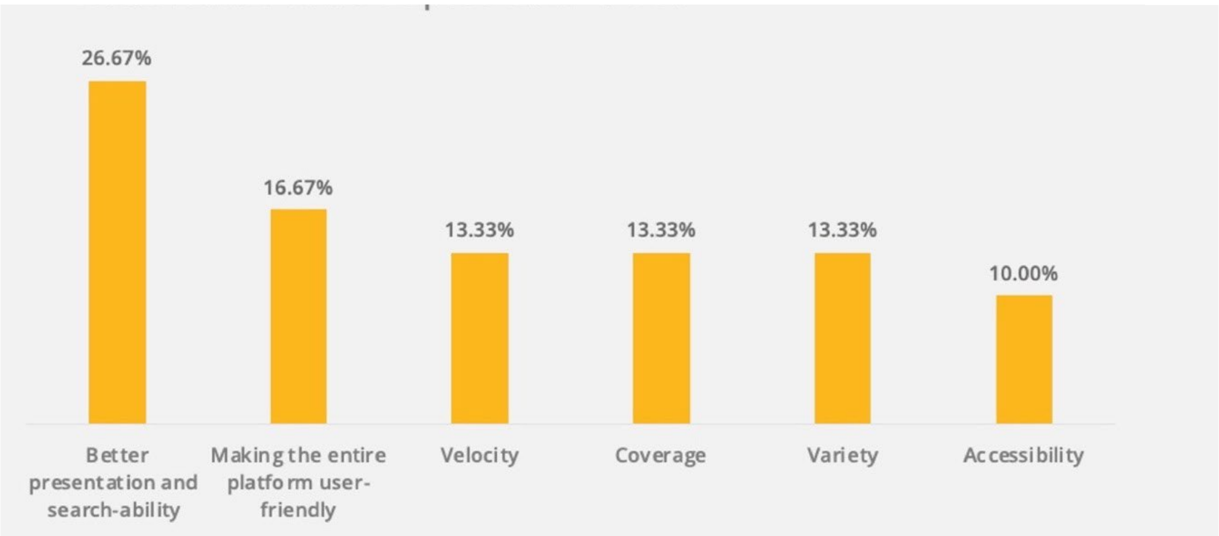
"APIs in some cases are sitting behind permissions, make the process longer"

The data on OGD

Curated list of findings & insights pertaining to the feedback on the data available on the OGD platform. Both the data audit by the taskforce and user feedback indicated similar gaps & pain points.

- High value of datasets is contextual to its end use
- Data credibility is questionable.
- Data quality triumphs over data quantity.
- Alternate sources have higher quality government data
- Lack of metadata standards
- Lack of information on data collection methodology
- Missing date and time context
- Catalog/dataset creation criteria unclear

- Poor data quality – inconsistency & incompleteness
- Unavailability of dataset, not continuous in coverage
- Missing historical dataset
- Missing multi-format dataset
- Adoption of standard vocabulary
- Inconsistent naming convention
- Blank files
- Lack of sync with source website



Platform and data issues that cause users to quit.

”Human and machine readable is important”

”What you see here is the dump and not data curation, no one has given a thought about the end users’ need”

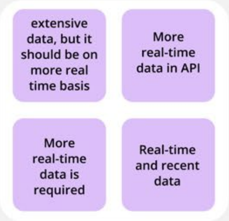
”Right now data is one pdf with all information together, one has to scrape to find what they are looking for”

”How would you put value on data? Its contextual”

Users’ view on volume-velocity-variety of OGD data

40%

of regular users stressed on the importance of having real-time data



extensive data, but it should be on more real time basis

More real-time data in API

More real-time data is required

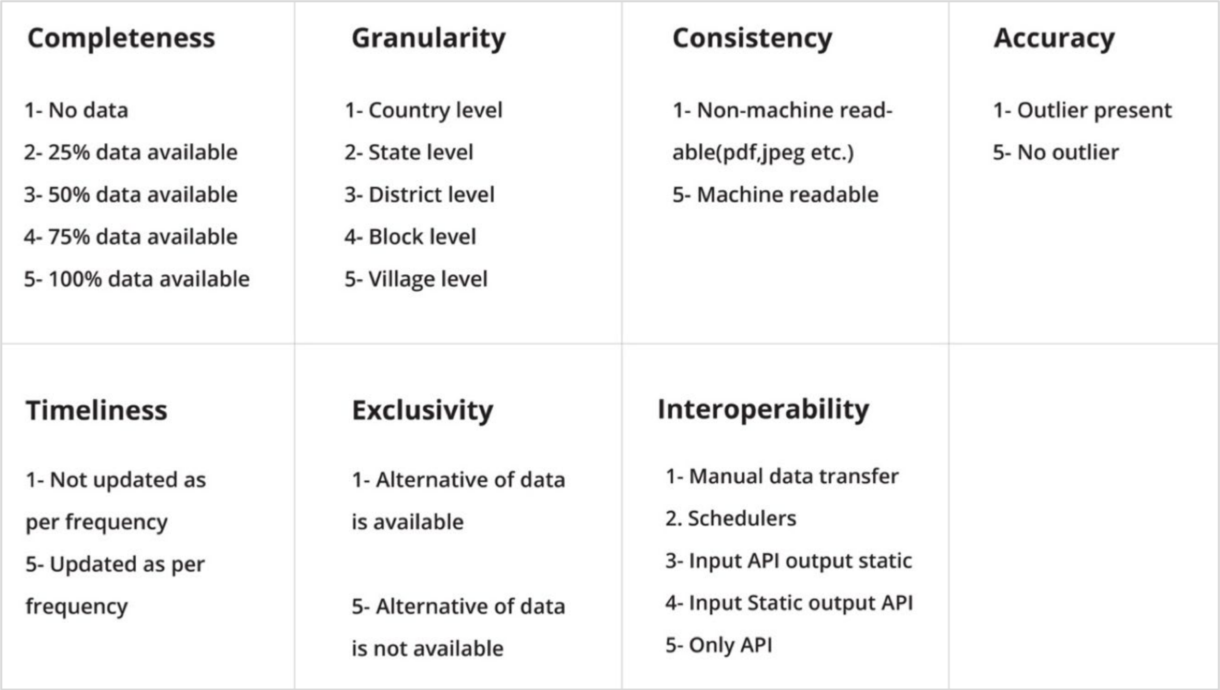
Real-time and recent data

50%

of users believe updating content frequently will improve the platform

HVD evaluation across the globe

Here are the findings from some of the important features and the corresponding scales after having evaluated several HVD frameworks across the globe.



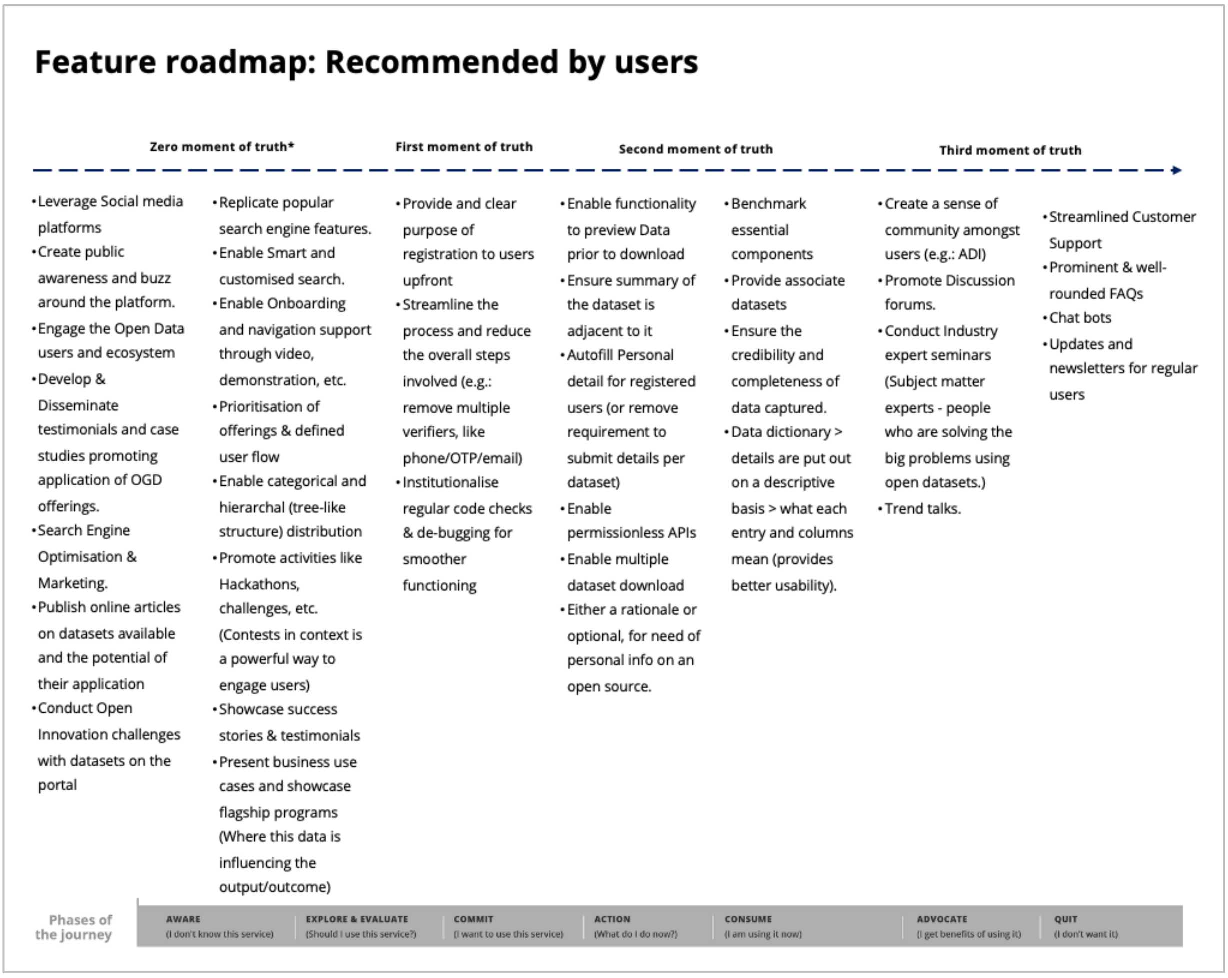
Completeness	Granularity	Consistency	Accuracy
1- No data 2- 25% data available 3- 50% data available 4- 75% data available 5- 100% data available	1- Country level 2- State level 3- District level 4- Block level 5- Village level	1- Non-machine readable(pdf,jpeg etc.) 5- Machine readable	1- Outlier present 5- No outlier
Timeliness	**Exclusivity**	**Interoperability**	
1- Not updated as per frequency 5- Updated as per frequency	1- Alternative of data is available 5- Alternative of data is not available	1- Manual data transfer 2. Schedulers 3- Input API output static 4- Input Static output API 5- Only API	

Data audit outcomes for the identified 19 HVDs

The data audit outputs helped understand and validate the end users' feedback and pain points. It involved the comprehensive audit of 19 unique, high value datasets.

User inputs on solution directions

Suggestions were given by various users, during primary research, that are opportunities to remedy the issues and shortcomings they faced while interacting and utilizing the data on the OGD platform. These inputs range from data correction to collection, to better parameters for high value data set identification and feedback on how best to redesign the website.



Feature roadmap: Recommended by users

Zero moment of truth* | First moment of truth | Second moment of truth | Third moment of truth

- Leverage Social media platforms
- Create public awareness and buzz around the platform.
- Engage the Open Data users and ecosystem
- Develop & Disseminate testimonials and case studies promoting application of OGD offerings.
- Search Engine Optimisation & Marketing.
- Publish online articles on datasets available and the potential of their application
- Conduct Open Innovation challenges with datasets on the portal

- Replicate popular search engine features.
- Enable Smart and customised search.
- Enable Onboarding and navigation support through video, demonstration, etc.
- Prioritisation of offerings & defined user flow
- Enable categorical and hierarchal (tree-like structure) distribution
- Promote activities like Hackathons, challenges, etc. (Contests in context is a powerful way to engage users)
- Showcase success stories & testimonials
- Present business use cases and showcase flagship programs (Where this data is influencing the output/outcome)

- Provide and clear purpose of registration to users upfront
- Streamline the process and reduce the overall steps involved (e.g.: remove multiple verifiers, like phone/OTP/email)
- Institutionalise regular code checks & de-bugging for smoother functioning

- Enable functionality to preview Data prior to download
- Ensure summary of the dataset is adjacent to it
- Autofill Personal detail for registered users (or remove requirement to submit details per dataset)
- Enable permissionless APIs
- Enable multiple dataset download
- Either a rationale or optional, for need of personal info on an open source.

- Benchmark essential components
- Provide associate datasets
- Ensure the credibility and completeness of data captured.
- Data dictionary > details are put out on a descriptive basis > what each entry and columns mean (provides better usability).

- Create a sense of community amongst users (e.g.: ADI)
- Promote Discussion forums.
- Conduct Industry expert seminars (Subject matter experts - people who are solving the big problems using open datasets.)
- Trend talks.

- Streamlined Customer Support
- Prominent & well-rounded FAQs
- Chat bots
- Updates and newsletters for regular users

Phases of the journey

AWARE	EXPLORE & EVALUATE	COMMIT	ACTION	CONSUME	ADVOCATE	QUIT
(I don't know this service)	(Should I use this service?)	(I want to use this service)	(What do I do now?)	(I am using it now)	(I get benefits of using it)	(I don't want it)

User recommendations

05 Recommendations & roadmap

Our recommendations can be classified into two broad categories:
1) Increase HVDs on OGD
2) Drive adoption of platform

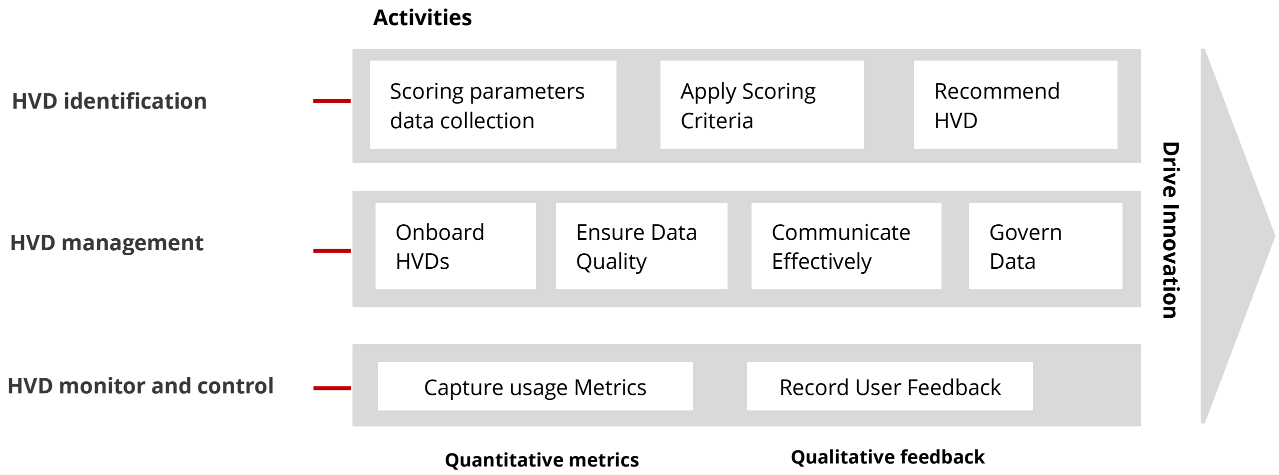
1. Increase HVDs on OGD

Discovering high value dataset (HVDs)

We recommend Chief Data Officers (CDOs) adopt the HVD framework shared in the report to identify, manage & monitor high value data sets for driving innovation. The framework identifies HVDs based on an integrated scoring criterion based of potential use cases, market value-based parameters, potential usage, etc. We further recommend the development of an AI/ML system that will auto update the input parameters into such a scoring mechanism. The framework also includes recommendations on regularly maintaining the quality of HVD list via onboarding, ensuring data quality, communicating updates to users under a robust mechanism for data governance.

HVD framework: The following were considered for a centralized framework to identify, manage & monitor high value data sets for driving innovation:

- HVD identification – This section focuses on identifying and classifying the data sets as HVDs or non HVDs.
- HVD management – This section includes activities such as onboarding HVDs, ensuring data quality, communicating effectively and governing data on a regular basis.
- HVD monitor & control – This section includes both quantitative metrics i.e., capturing usage metrics and qualitative inputs I.e., recording user feedback to ensure sustained usefulness and relevance of datasets tagged as HVDs.



HVD identification – areas of evaluation:

Parameters have been divided into three categories for evaluation purposes.

Category 1: Use case based **OR** **Category 2: Usage based** **OR** **Category 3: Market value based**

Category 1- Use case based

Based on the afore mentioned approach, we have listed few illustrative use cases, for which the corresponding datasets that can be qualified as HVD, and likewise can be updated or curated.

Sector	Datasets
Agriculture	Data for assessing global warming implications on weather/seasons for better planning
	Data for predicting and mitigating crop diseases
	Data for predicting rainfall & other seasonal occurrences
	Data for assessing soil quality
	Precision farming
	GIS backed data for better planning
	Geographical condition overall
Research & academic	Data for PhDs, academic research for papers, capstone/industry projects
	Data for think tanks/incubators within colleges
	Data for microeconomics
Retail	Geo tagged, pin code level data for logistics
	Data to identify possible markets from income distribution data
	GIS backed data to identify area size, location for store/warehouse
	GIS backed data to identify distances (shortest/farthest etc.) From stores
Health	Data for Covid mortality tracking, district level
	Data for connecting midwives to pregnant women
	Data for pin code level, geo tagging for policy makers
	Cancer data (across types) for understanding severity, fatality, region, age etc. for cancer management & care
	Data for death and overall impact for healthcare programs/facilities/resource management
	Data for illness by region for healthcare programs/facilities/resource management
Insurance	Data for rainfall prediction for agricultural insurances
	Data for pin code level, geo tagging for insurance providers
	Data for mortality rate based on pin code wise
	Data for Illness by region
Government & official bodies	Data for government healthcare systems to design their policies
	Data for geocode tenders related to waterworks
	Data for justice and law: communities related to prison reform, policing, courts etc
	GIS backed data for planning
Urban development	Geographic information system (GIS) backed data to solve for urban & civic issues
	Data for urban area wise flooding data (low lying areas, repeated flooding etc.)
	Data for connecting public transports to inaccessible locations
	Data for identifying sites for train tracks

	Data for urban planning & Infrastructure
	Pin code level data
Forest department	Data for identifying animal human conflict zones
	Data for tracking forest fires
	Data for identifying forest covers within urban areas
	Data for tracking animal movement, like elephant paths, augmenting the past/old data to latest
	Geo tagging data
Sustainability	Air quality index data to assess impact and strategize measures accordingly
	Geographical condition overall
	Pollution data to assess impact

Category 2- User demand-based Assessment

User experience metrics helps in assessing if the dataset is of high user demand or not. This can be assessed through below mentioned dimensions:

- Number of downloads- Ensure the number of downloads (APIs, direct downloads, and visualizations) in the last 1 year is greater than the above agreed threshold.
- Rating- Ensure the rating of datasets is greater than 4 and is also rated by the above agreed threshold.
- Comments- Ensure the number of comments posted by the users is greater than the above agreed threshold.
- Shares- Ensure the number of shares by users is greater than the above agreed threshold.

The rating would be given to each of the above-mentioned dimensions in the form of ‘Yes’ or ‘No’ based on the criteria met.

Category 3- Market value based

Market value based is derived based on the below categories:

- Market parameters
- Degree of impact
- Relevance

The brief description of the said categories is mentioned in detail below:

- **Market parameters** are a qualitative measurement that help us understand if the dataset has implications in any of the areas below:

- Social conditions e.g., weather data, public housing data, etc.
- Economic impact e.g., anticipated growth in income. Projected saving of public funds, etc.
- Environmental impact e.g., water & air quality
- State of public services e.g., location-based data, tourism data

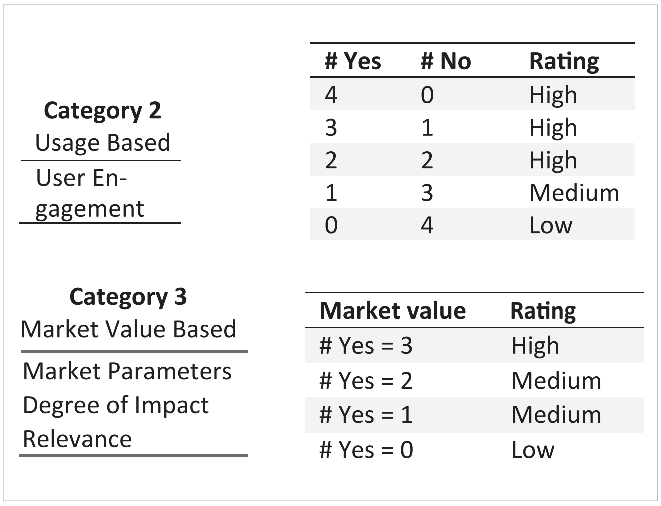
- **Degree of impact** will help us identify whether the dataset can cause a significant impact in social, economic, environmental, political, legal or technological factors

- **Relevance** also plays a key role in identifying whether the dataset is relevant either "now" or "in near future" or whether the dataset is in high demand by the community.

The rating would be given to each of the above-mentioned dimensions (i.e., market parameters, degree of Impact and relevance) in the form of 'Yes' or 'No' based on the criteria met.

Category 1 - Use case based: If the dataset falls into one of the use cases listed then the dataset is HVD

Else

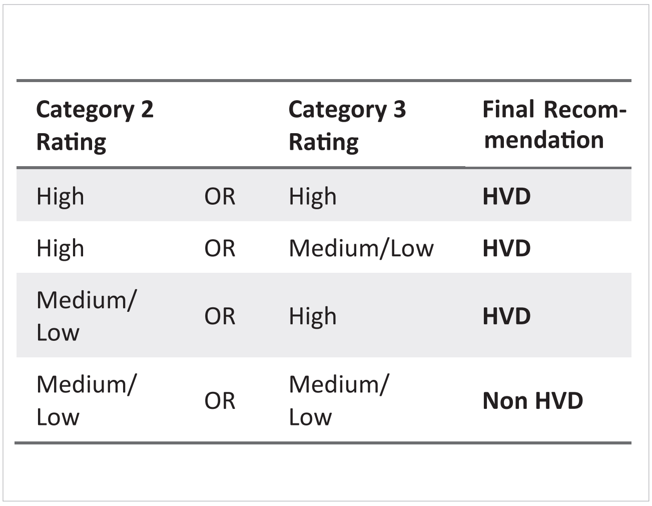


Category 2
Usage Based
User Engagement

# Yes	# No	Rating
4	0	High
3	1	High
2	2	High
1	3	Medium
0	4	Low

Category 3
Market Value Based
Market Parameters
Degree of Impact
Relevance

Market value	Rating
# Yes = 3	High
# Yes = 2	Medium
# Yes = 1	Medium
# Yes = 0	Low

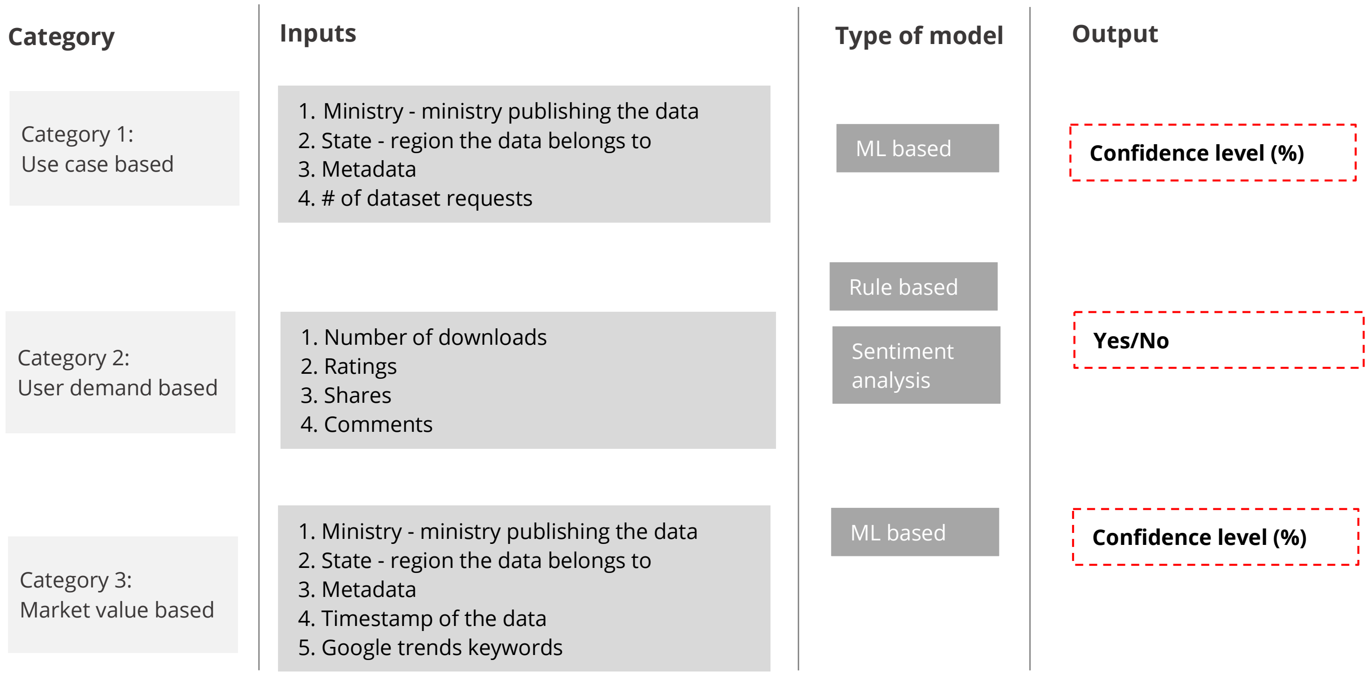


Category 2 Rating		Category 3 Rating	Final Recommendation
High	OR	High	**HVD**
High	OR	Medium/Low	**HVD**
Medium/ Low	OR	High	**HVD**
Medium/ Low	OR	Medium/ Low	**Non HVD**

HVD identification – scoring criteria: for bifurcating the datasets into HVD, we need to sum the number of yes and nos from areas of evaluation and map the same in the respective row in the matrix below for final recommendation.

Based on the above scoring mechanism, we understand that for use case-based approach (category 1), if any of the datasets fall under the identified use cases, they will qualify as an HVD. Further, any datasets passed through the categories i.e., usage based (category-2), or market value based (category-3) will qualify as HVD in case the rating is 'high.'

Automating the framework - Data points for areas of evaluation: As the above HVD management framework is a manual process, to automate the same, we have come up with the framework that can be used to simplify the process for HVD identification. The proposed framework is the blueprint of the automation framework and needs to be further developed as per the OGD roadmap.



Category	Inputs	Type of model	Output
Category 1: Use case based	1. Ministry - ministry publishing the data 2. State - region the data belongs to 3. Metadata 4. # of dataset requests	ML based	**Confidence level (%)**
Category 2: User demand based	1. Number of downloads 2. Ratings 3. Shares 4. Comments	Rule based Sentiment analysis	**Yes/No**
Category 3: Market value based	1. Ministry - ministry publishing the data 2. State - region the data belongs to 3. Metadata 4. Timestamp of the data 5. Google trends keywords	ML based	**Confidence level (%)**

For the above-mentioned categories, we have also identified the different type of dataset inputs that can be input to ML/AI algorithm and predict the confidence level or threshold level to categorize the dataset as HVD

The brief description about each input for respective categories are mentioned below:

Category 1 (Use case based)

Certain ML models with following inputs related to the data set can help predict the confidence level of the data set being an HVD

1. Ministry – Certain ministries such as Health Ministry will address high value use cases hence the data sets from these ministries will mostly get qualified as HVD.

2. State – Certain regions in combination with other parameters can help identify high value use cases hence HVDs.
3. Metadata – The metadata collated from the datasets will be helpful for the identification of data sets to be classified as HVDs.
4. Number of dataset requests – The number of requests raised for each data sets will be helpful in gauging whether the required use case is high in demand or not and based on the demand, the same could be classified into HVD.

Category 2 (User demand based)

1. Number of downloads- More the number of downloads, higher will be the probability to qualify as HVD.
2. Ratings- Higher the rating from the users, higher will be the probability to qualify as HVD.
3. Sharing - More the number of sharing, higher will be the probability to qualify as HVD.
4. Comments - More the number of comments shared on OGD platform and social media, higher will be the probability to qualify as HVD.

Category 3 (Market value based)

This model could be combined with the same ML model in Category 1 as the inputs have a huge overlap

1. Ministry – Certain ministries such as Health Ministry will address high value use cases hence the data sets from these ministries will mostly get qualified as HVD
2. State – Certain regions in combination with other parameters can help identify high value use cases hence HVDs
3. Metadata – The metadata collated from the inputs provided will be helpful for identifying use case as HVD
4. Time stamp of the data – This will help us in evaluating whether the dataset consists recent data or not
5. Google Trends keywords – This will help us in analyzing which type of data is gaining more relevance or importance in today’s time

Enriching existing HVDs

We recommend enriching 19 existing HVDs on the OGD platform (as mentioned in this report) and monitoring their quality regularly based on data quality parameters identified.

As part of this study, we have identified data quality gaps in 19 existing HVDs shared by MeitY in the areas of agriculture, census, shipping, healthcare, finance, and telecom.

These gaps pertain to completeness, consistency, timeliness, and relevancy of these datasets, that we feel are hampering user adoption by creating usability issues.

Sr #	Datasets	Completeness	Consistency	Timeliness	Relevancy
1	Current daily price of various commodities from various markets	✗	☑	☑	☑
2	Real time air quality index from various locations	✗	☑	✗	☑
3	Variety-wise daily market rrices of cotton	✗	☑	✗	☑
4	Village amenities, Census 2011*	☑	☑	☑	☑
5	Town amenities, Census 2011*	☑	☑	☑	☑
6	Abstract of receipt from 2019-20 to 2022-22	☑	☑	☑	☑
7	Growth of Indian shipping	☑	☑	☑	☑
8	NIN health facilities with geo code and additional parameters (updated till last month)	✗	☑	☑	✗
9	Voice call quality customer experience (MY-CALL app)	✗	✗	☑	✗
10	Farmers queries in kisan call centre (kcc)	✗	✗	✗	✗
11	All India crowdsourced mobile data speed measurement	✗	☑	☑	✗
12	Gender budget from 2019-20 to 2022-22	☑	☑	☑	☑
13	Tax revenue from 2019-20 to 2022-22	☑	☑	☑	☑
14	Allocations for the welfare of schedule caste	☑	☑	☑	☑

15	Central sector scheme from 2019-20 to 2022-22	☑	☑	☑	☑
16	His indicator-wise monthly	☑	☑	✗	☑
17	Voice call quality customer experience	✗	✗	☑	✗
18	Company master data	☑	☑	☑	☑
19	List of msme registered units under udyam	✗	✗	✗	✗

Augment HVD list

We recommend augmenting the HVD list by applying the HVD framework recommended as part of this report on the 93 datasets identified during the study as valuable to users.

This list was synthesized based on extensive user research, secondary research, data audit and survey. We recommend CDOs in each ministry to find a comparable dataset through internal discovery and uploading them on OGD in case the HVD Framework categorizes them as high value. *(Refer to the appendix for the detailed version)*

HVD management guidance: Other important enabling the facilitation of access and creation of robust management of data are:

Data Governance (DG)

S. #	Particulars	Description
1	Governance structure	An effective governance structure will allow creation of value, through innovation and development, provide accountability and control systems commensurate with the risks involved
2	DG processes	DG processes will allow to manage the availability, usability, integrity and security of the data based on internal data standards and policies that also oversee data access
3	Policies	A robust data governance policy will help to consistently address the developments and scenarios that may arise related to its creation, processing, use, and sharing of data for innovation
4	Data dictionary	A centralized repository should be maintained wherein information about data such as meaning, relationships to other data, origin, usage, and format can be safely stored and accessed
5	Interoperability	Defining interoperability and tools to help operationalize it will help in sharing and making use of information more widely
6	Dataset prioritization	Proper mechanism should be in place to ensure the prioritization of datasets available on the platform based on the intrinsic value of data
7	Data privacy	Data privacy rules should be incorporated to ensure the data shared by customers is only used for its intended purpose
8	Data security	Defining data security will require a roadmap to protect digital data from actions of unauthorized users and protection from cyber attacks

Ensuring data quality

S. #	Particulars	Description
1	Data quality rules	Data quality rules should be applied on regular basis to validate data values and records
2	Data profiling	Data profiling will help to understand the structure of data, relationships between data sets, and how it could potentially be used most effectively
3	Completeness	Data should be complete in all respects. There should be no gaps or missing information for data to be truly complete.
4	Consistency	Data consistency should be maintained across database to ensure that data kept at different places should match.
5	Auditability	Auditing will allow to review the datasets to assess the quality or utility
6	Accuracy	Data should always be accurate and updated. Data values should store correct value and must be represented in a consistent and unambiguous form
7	Cleanliness	Cleanliness in data will help in detecting and correcting (or removing) corrupt or inaccurate data in the datasets
8	Clarity	Data set should have clarity (i.e., it should include appropriate metadata, illustrations such as graphs and maps, limitations etc.)
9	Integrity	Data integrity will help in creating meaningful and valuable datasets
10	Validity	Capabilities that enable data should be validated in line to the project requirements

Communicate effectively

S. #	Particulars	Description
1	E-mail communication	Proper facility should be available to transfer the data via email or other modes.
2	Training	Proper onboarding and training sessions must be organized at regular intervals to all the stakeholders involved
3	Marketing campaigns	Proper campaigns should be planned and held to promote awareness on the data available on the platform
4	Subscription management	Subscription management should be in place to ensure proper management of users account by making sure that subscribed members notifications, updates and other additional benefits on a regular basis

Based on the above HVD management guidance, we understand that once the datasets are published, we need to fulfill the quality standards not only from the data perspective but also w.r.t right format, ensuring proper metadata to be in place by reviewing and approving the same on regular basis, making sure that said metadata is easily recognizable by users to enable the right prioritization of datasets.

We also need to ensure proper communication via e-mails, proper training material to be published along with datasets.

HVD monitor & control: HVDs, once published should be frequently monitored based on various quantitative and qualitative inputs. This helps in continually improving the quality of HVDs. The metrics for continuous monitoring and evaluation of HVDs include:

- **Download frequency-** Based on the number of downloads, metrics can be shared to HVD publishing entities
- **Number of views-** Number of views and % of downloads will play a role in understanding datasets of interest that were viewed but not downloaded.
- **Rating/User review**- Incorporating number of reviews/ and positive negative reviews will give us quantitative feedback of datasets uploaded.
- **User feedback**- User feedback based on reviews, comments, suggestions and requests should be incorporated into the datasets.
- **Use-case adoption-** Helps clearly identify areas of interest to end users. Also helps in creating use cases that are not discovered yet.

The above metrics/KPIs may be leveraged to provide proper visibility to CDOs/data contributors w.r.t how datasets are being used. Further, the above metrics and inputs should be mapped against each dataset/HVD and the same shall be monitored over a dashboard using a visualization tool.

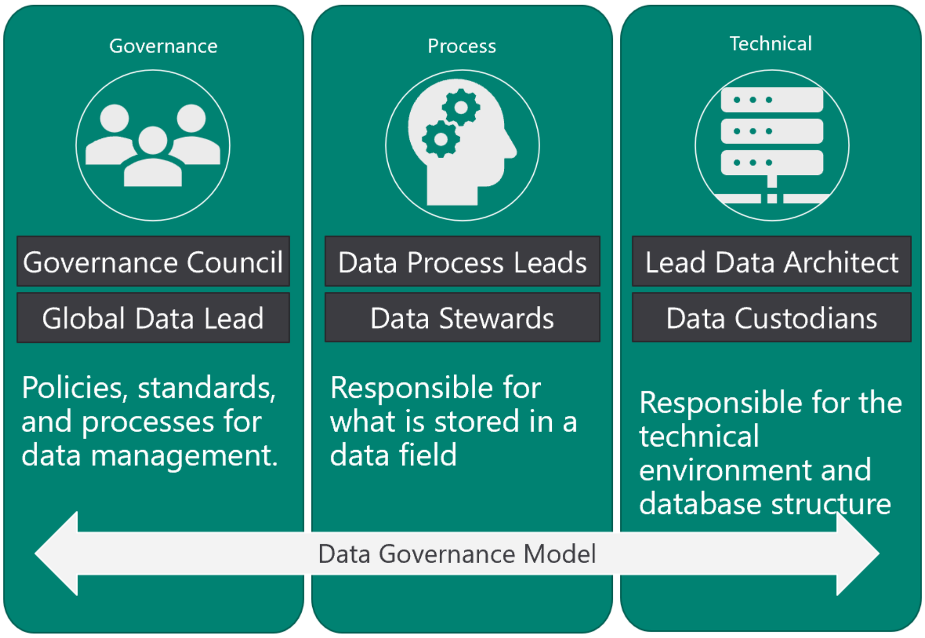
Data governance

Data governance is one of the most important considerations for harnessing the value of data and therefore for the OGD architecture. The primary objective is to enable consistent data access across the user community irrespective of location, speed, and structure of data. Data governance covers processes and tools to streamline data quality, accuracy, interoperability, access, and use. Ideally, the Data architecture pattern would be with automated data management & integration should leverage AI/ML in augmented data catalog, knowledge graph, semantics, and data quality management. For example, governance tools utilizing machine learning capabilities to automate the most tedious tasks involved in data catalogue creation, that is data discovery, ingestion and enrichment of metadata, as well as identification of relationships between metadata ensure that latest and most robust data is available to those who need it. Data consumers can find, understand, and use relevant datasets more quickly – reducing time to insights and improving business decisions.

Baking the data governance strategy in with the overall OGD modernization strategy is essential to ensure that the necessary Governance capabilities can be embedded and appended over time. A proper data governance strategy should address –

- Information strategy & architecture - Data governance and analytics governance are well-defined frameworks integrated with each other. The key architectural perspectives are-
  - People & organization (organization models, PM office etc.)
  - Processes, policies & compliances (data ops, service model, communication model, etc.)
  - Technology management (reference architecture, infrastructure mgmt. etc.)
- Metadata / data catalog / master data management by defining business glossary and performing the metadata integration with overall data governance operations. This includes data control (quality, transparency, auditability, security)
- AI/ML driven intelligent data quality by defining the data quality approach and performing necessary integration of data quality tools with source and destination systems. E.g., augmented structured DQM, textual DQM, audio/video DQM etc.

One critical governance aspect (other than data governance) for successful OGD modernization is competency management. This stands out as one single point of focus to ensure the availability on relevant quality datasets in a timely manner. An organization structure that has the right people in the right roles is crucial to achieving organizational goals. Regarding data management, at a broad level it is important to have well defined set of roles taking care of governance, process and technical tasks as detailed in the below visual.


Governance
Process
Technical
Governance Council
Global Data Lead
Data Process Leads
Data Stewards
Lead Data Architect
Data Custodians
Policies, standards, and processes for data management.
Responsible for what is stored in a data field
Responsible for the technical environment and database structure
Data Governance Model

To ensure ministries/departments are able to publish quality datasets consistently, proper training needs to be conducted on regular basis for each of the roles depicted above.

Data management:

Key recommendations pertaining to data management, specifically with regards to data storage, data quality, metadata management, and ontology management are given below:

- Data fabric – Capability for a data fabric to attain flexible, reusable, and augmented data integration pipelines utilizing knowledge graphs, semantics, and ML/AI on active metadata. The fabric enables less technical business users/ subject matter experts to quickly find, access, integrate and share data.

- Polyglot persistence & knowledge graphs: Use specialized databases (both NoSQL and relational) for different purposes within the same web application. One-database-fits-all may not be the de facto choice. Polyglot persistence will bring in choice – choice to select best data store for a given data type or purpose. Polyglot persistence will leverage the strength of multiple data stores. additionally, knowledge graph will provide a knowledge base that will lie on top of polyglot persistence and use a graph-structured data model or topology to integrate data.

- Microservices & API-fiction: Analytical applications will be developed in an accelerated fashion using data and analytics assets, microservices, and APIs.These will be exposed as data/analytics products. The architectural pattern in AI applications will be the use of scalable algo services and business services using microservices and APIs at right grain. This will enable multi-level scalability of “as a product”, but governance needs to be carefully crafted in the architecture to keep complexity under control.

- Data storage mechanism: Unified modern data store with extreme scalability for all available data. The system should be robust enough to cater to structure as well as unstructured data.

- Data ingestion / real time open data: Access to content from different heterogeneous sources though a single access point with a smooth, near real time data ingestion technology can be put in place - includes automated extraction of content from different institutional back-ends to incorporate them, in an automated way, to an open data catalog.

- Augmented data quality management (DQM): This focus areas will include use of AI/ML to augment the DQM capabilities. This will include both, employment of AI/ML for DQM on structured data, and extension of DQM into the unstructured realm, beyond text and into the image, video & audio datasets, as detailed below.

– Augmented structured DQM: Pattern/rule recommendation based on semantic mapping; Anomaly detection based on historical patterns and on context of the data set; Rules usage pattern & adaptive Rules to improvise DQ process efficiency, business case specific metrics.

– Textual DQM: NLP metrics e.g., perplexity, BLEU; Measuring conformance to templates e.g., for conformance of legal contracts and completeness of field inspection notes; anomaly detection between documents submitted for KYC.

– Image & video DQM: Computer vision metrics, quality sensing e.g., to establish suitability of images taken as part of the sanitation inspection process; categorization & classification e.g., to establish suitability for intake into information filtering systems; native extraction etc.

– Audio DQM: Para-linguistic quality parameters in terms of Jitter, Shimmer, HNR.

- Augmented Metadata Management (MM) / Semantics - Like DQM, AI/ML will be a critical enabler of next gen MM which is pertinent with OGD management.

– Discovery – Fuzzy matching BE's and PDEs using technique such as Levenshtein Distancing; Natural Language interrogation e.g., NL querying by APIs' and conversational interaction.

– Lineage management - System to system; code interpretation e.g., parsing engine for data mapping and bus. rules embedded in code; physical system to semantic layer; functional summarization e.g., noise reduction on & summarization of physical lineage to generate technical & functional lineage.

– Dataset level metadata – Auto profiling at the dataset level; semantic mapping of whole datasets; convention recommendation e.g., treatment rule set recommendation based on the context of the data set and analytic use case.

- Document level metadata – Metadata based auto categorization of documents, native extraction of items of interest e.g. customer, product for document classification.

- Data annotation – Metadata and creation of data to build training dataset to feed into AI Models as well as have explain-ability for the learning/meta-learning process and auditability.

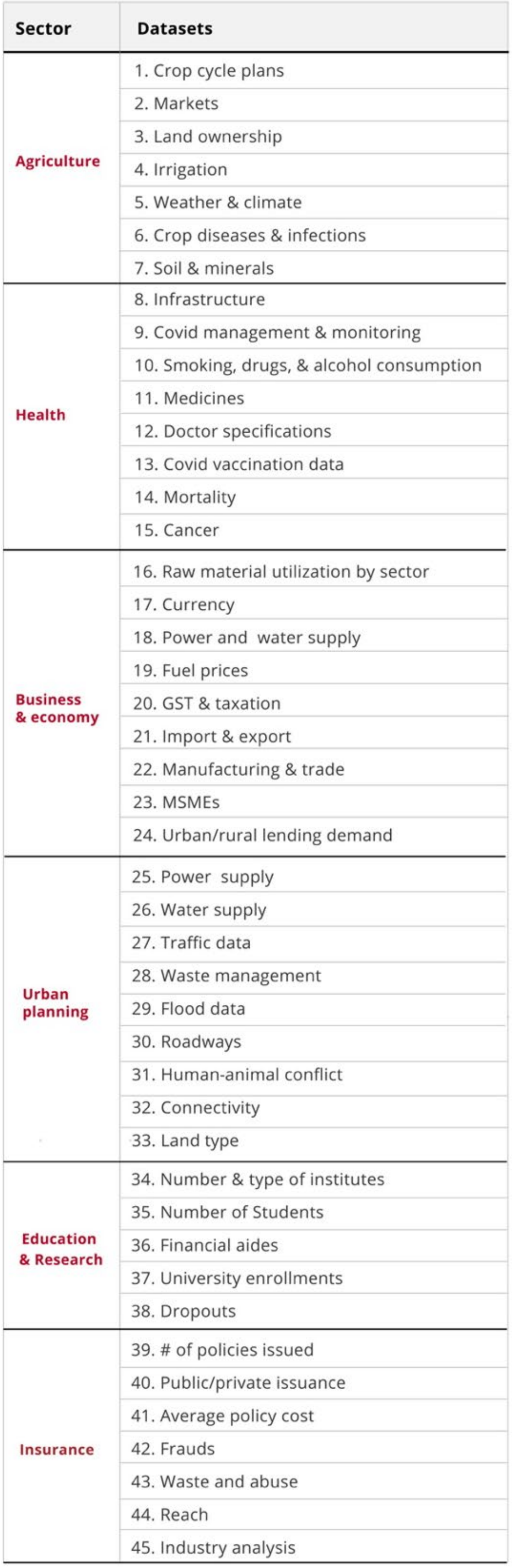
- **Ontology management** – This can be a key technology enabling semantic interoperability and integration of data and processes.

- Ontology recommendation – Taxonomy e.g., auto classification of entities; normalization of hierarchies, tiers & facets e.g., building part specifications from documents; auto maintenance e.g., detection & application of changes required to ontological structures based on new data; hierarchy deduction e.g., building hierarchy of parts for aircraft manufacturing or oil rig assets.

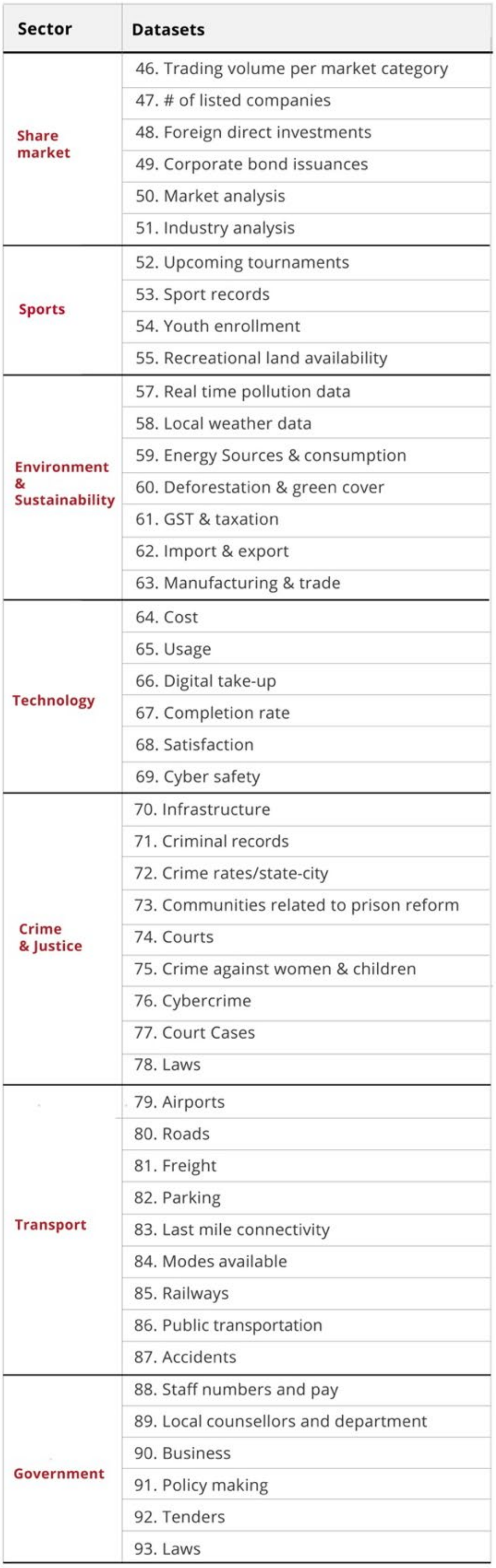
- Ontology based data management – Governance e.g., case mgmt. and attestation; supporting OBDM (Ontology Based Data Mgmt.) and data exchanges.

Data-centric engagement with the community:

- At the dataset level, discussions between consumers and providers should be enabled for quick clarifications, feedback and sharing ideas.
- The portal should host a discussion forum to allow open conversations among consumers and with producers on various aspects of the portal and datasets.
- Use of an integrated feedback gathering mechanism such as NIC's RAS from users who download datasets or use APIs.
- A unified view of resources/datasets to allow users to explore all information about a specific catalog/dataset, including and not limited to community feedback and contributions, potential for reuse, details about collection methodology and metadata, etc.
- A public facing detailed time series dashboard of demand, availability and re-use patterns at the dataset, catalog and publisher levels.
- Allow extraction of the full package list (list of all resources/ catalogs and associated metadata) programmatically.



Sector	Datasets
Agriculture	1. Crop cycle plans
	2. Markets
	3. Land ownership
	4. Irrigation
	5. Weather & climate
	6. Crop diseases & infections
	7. Soil & minerals
Health	8. Infrastructure
	9. Covid management & monitoring
	10. Smoking, drugs, & alcohol consumption
	11. Medicines
	12. Doctor specifications
	13. Covid vaccination data
	14. Mortality
	15. Cancer
Business & economy	16. Raw material utilization by sector
	17. Currency
	18. Power and water supply
	19. Fuel prices
	20. GST & taxation
	21. Import & export
	22. Manufacturing & trade
	23. MSMEs
	24. Urban/rural lending demand
Urban planning	25. Power supply
	26. Water supply
	27. Traffic data
	28. Waste management
	29. Flood data
	30. Roadways
	31. Human-animal conflict
	32. Connectivity
	33. Land type
Education & Research	34. Number & type of institutes
	35. Number of Students
	36. Financial aides
	37. University enrollments
	38. Dropouts
Insurance	39. # of policies issued
	40. Public/private issuance
	41. Average policy cost
	42. Frauds
	43. Waste and abuse
	44. Reach
	45. Industry analysis



Sector	Datasets
Share market	46. Trading volume per market category
	47. # of listed companies
	48. Foreign direct investments
	49. Corporate bond issuances
	50. Market analysis
	51. Industry analysis
Sports	52. Upcoming tournaments
	53. Sport records
	54. Youth enrollment
	55. Recreational land availability
Environment & Sustainability	57. Real time pollution data
	58. Local weather data
	59. Energy Sources & consumption
	60. Deforestation & green cover
	61. GST & taxation
	62. Import & export
	63. Manufacturing & trade
Technology	64. Cost
	65. Usage
	66. Digital take-up
	67. Completion rate
	68. Satisfaction
	69. Cyber safety
Crime & Justice	70. Infrastructure
	71. Criminal records
	72. Crime rates/state-city
	73. Communities related to prison reform
	74. Courts
	75. Crime against women & children
	76. Cybercrime
	77. Court Cases
	78. Laws
Transport	79. Airports
	80. Roads
	81. Freight
	82. Parking
	83. Last mile connectivity
	84. Modes available
	85. Railways
	86. Public transportation
	87. Accidents
Government	88. Staff numbers and pay
	89. Local counsellors and department
	90. Business
	91. Policy making
	92. Tenders
	93. Laws

Privacy preservation

In the context of open data, privacy preservation technologies allow data officers to share sensitive datasets that may be leveraged for innovation and research, while preserving the privacy, security, and trust of citizens.

Access to privacy preserving technologies on the OGD platform may lead to an increase in the ease of sharing sensitive data as CDOs leverage these technologies to address various privacy risks ultimately leading to an augmented list of high-quality datasets.

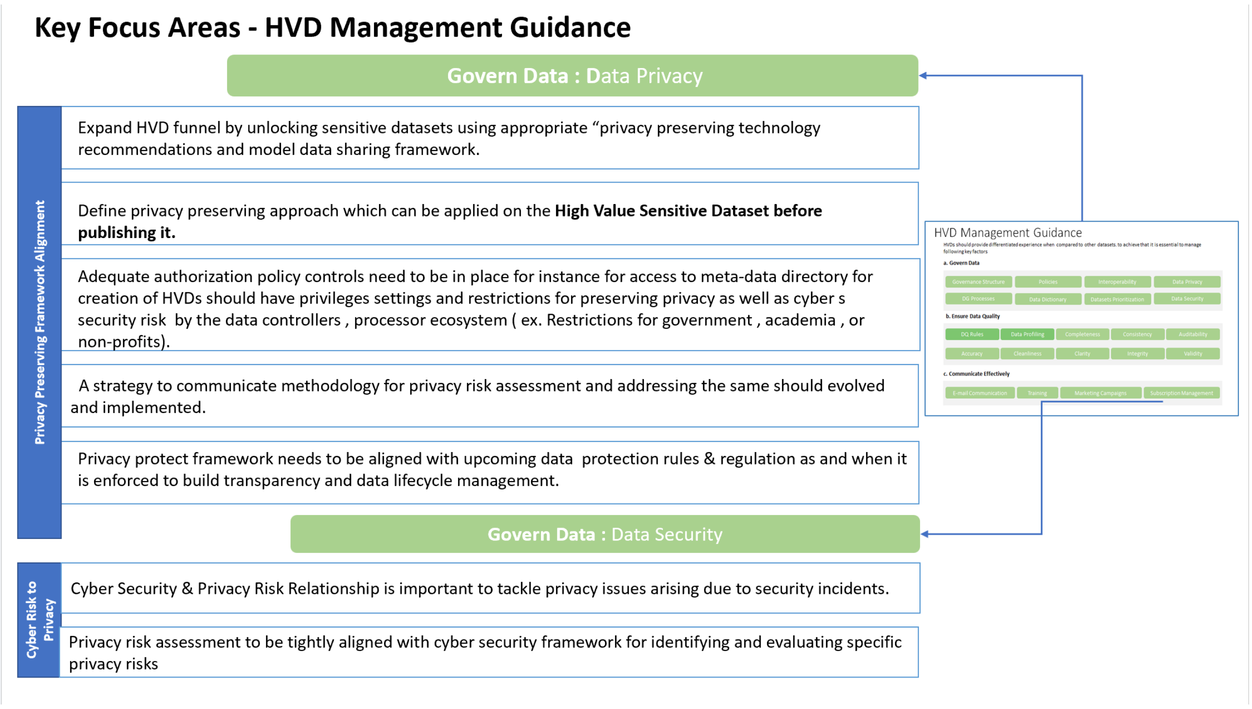
This report recommends a privacy preserving framework that assures privacy in best possible way without any data distortion and ambiguity. This is done through adoption of technology best suited to address the identified risk in adherence to the governance policies specified at the macro level.

It is further recommended that the Privacy Preserving Framework be integrated with the Cyber Security Framework within the overall Data Exchange Ecosystem to manage privacy risks arising from data processing. Managing cyber security contributes towards managing privacy risks via handling incidents related to loss of confidentiality, integrity, or availability of HVDs and thus alleviating the resultant damage to reputation or more tangible harm such as loss of economic or social opportunity. However, Privacy incidents can arise due to inadequate handling of personal or sensitive information which is defined as follows as per VCDPA:

'Personal data' is defined as any information that is linked or reasonably linkable to an identified or identifiable natural person but does not include de-identified data or publicly available information. The VCDPA's definition of 'personal data' roughly aligns with 'personal data' under GDPR.

'Sensitive data' is a category of 'personal data' that includes the following:

- Personal data revealing racial or ethnic origin, religious beliefs, mental or physical health diagnosis, sexual orientation, citizenship, or immigrations status;
- The processing of genetic or biometric data for the purposes of uniquely identifying a natural person;
- The personal data collected from a known child; or
- Precise geolocation data.



Key Focus Areas - HVD Management Guidance

Govern Data : Data Privacy

Privacy Preserving Framework Alignment

Expand HVD funnel by unlocking sensitive datasets using appropriate “privacy preserving technology recommendations and model data sharing framework.

Define privacy preserving approach which can be applied on the **High Value Sensitive Dataset before publishing it.**

Adequate authorization policy controls need to be in place for instance for access to meta-data directory for creation of HVDs should have privileges settings and restrictions for preserving privacy as well as cyber s security risk by the data controllers , processor ecosystem (ex. Restrictions for government , academia , or non-profits).

A strategy to communicate methodology for privacy risk assessment and addressing the same should evolved and implemented.

Privacy protect framework needs to be aligned with upcoming data protection rules & regulation as and when it is enforced to build transparency and data lifecycle management.

Govern Data : Data Security

Cyber Risk to Privacy

Cyber Security & Privacy Risk Relationship is important to tackle privacy issues arising due to security incidents.

Privacy risk assessment to be tightly aligned with cyber security framework for identifying and evaluating specific privacy risks

HVD Management Guidance

HVDs should provide differentiated experience when compared to other datasets. to achieve that it is essential to manage following key factors

a. Govern Data

b. Ensure Data Quality

c. Communicate Effectively

Key recommendations for OGD Privacy Preserving Framework

The following privacy considerations observed through the study of global use cases and examples may be contextualized and applied to OGD:

- Privacy-preserving emerged in several case studies, demonstrating ways in which sound privacy practices can be used without undermining the opportunities for data sharing. The adoption should point towards harnessing the value of data for social good purposes and for improving the economic, social, and technological benefits to governments, society, and citizens.

- Promote the use of privacy-enhancing technologies to safeguard sensitive information in data sharing scenarios, for example
  - Limit observability and likability: HVD available on local device, data encrypted at rest, in transit, as well as while in use (confidential computing)
  - Limit Identification of Individuals (e.g., de-identification privacy techniques including anonymization, tokenization)
  - Limit the formulation of inferences about individuals’ behaviour or activities (e.g. data processing is decentralized, distributed architectures, differential privacy involving data distortion while maintaining key statistical properties along with enforcing a privacy budget)
  - Limit disclosure of data elements (e.g. authorization policies)
- Taking a privacy-by-design approach while implementing data privacy initiatives is another practice that can help mitigate privacy risks while developing

new data collaborations (releasing new HVDs on OGD platform). Framework may require governance model on complete data life cycle management - data collection, processing, storing, sharing, retention, deletion

- Open government data platforms in various countries follow international best practices, ISO standards (such as ISO/IEC 38505 for data governance, ISO27701 for privacy information management), and practical tools like the NIST cybersecurity framework. It is recommended to align with a broadly adopted and standardised framework like NIST Privacy Framework for application of key privacy risk management practices.

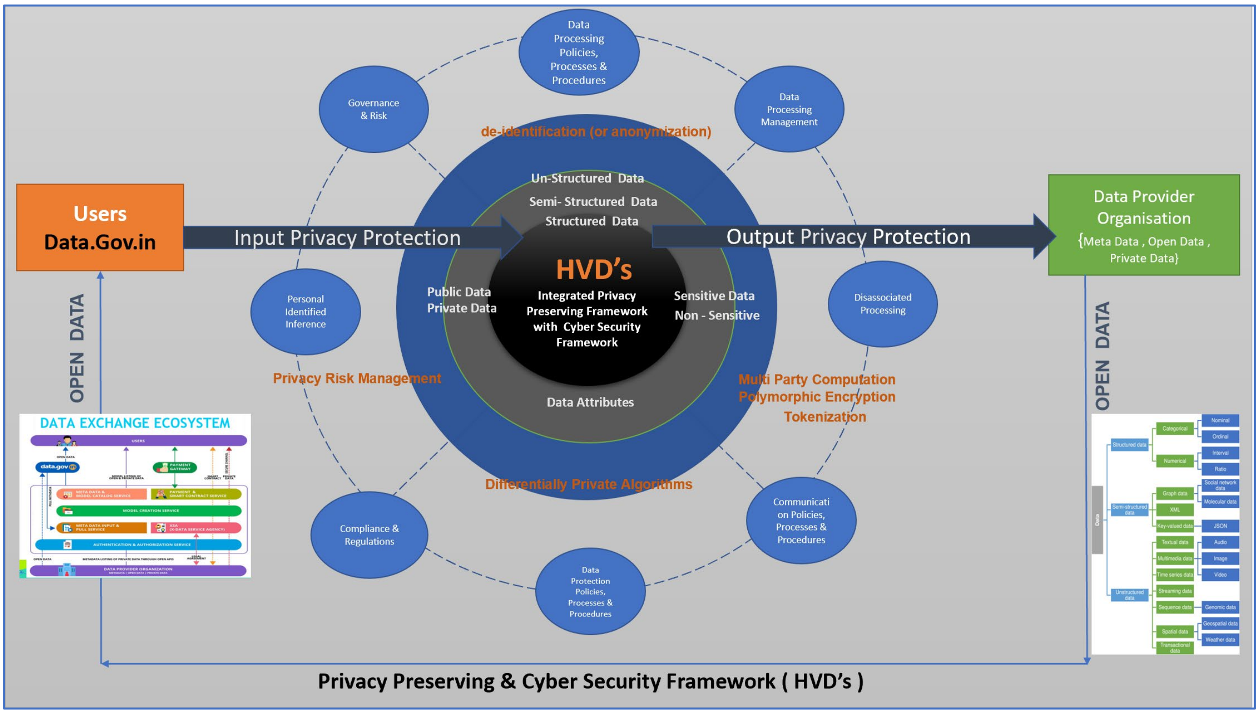
- Privacy risk assessment is a key process in privacy risk management for identifying and evaluating specific privacy risks. In general, privacy risk assessments produce the information that help CDOs weigh the benefits of sharing the HVDs in OGD platform against the associated privacy risks and determine the appropriate response—sometimes referred to as proportionality.

- The ways of responding to risk include:
    - Reducing the risk by taking measures that make it less likely or less severe (for example, by applying technical or policy measures to systems, products, or services)
    - Avoiding the risk by deciding that the risks outweigh the benefits and not sharing the HVD
    - Deciding that the risks are acceptable because they are not likely to happen or not likely to cause harm, or because the benefits of taking the risk outweigh the potential harms

- It is proposed to build a continuous learning mechanism for risk assessment of data sets for example, AI based methodology for data classification and ultimately sensitivity labelling based in input features like classification data, use case, usage and market value pertaining to data set, dept releasing the dataset etc. The features leading to an accurate sensitivity labelling will get refined over time, as is the case in any AI approach.

Privacy protection framework

The representation below shows the HVD privacy persevering framework and how the current data exchange ecosystem may be integrated with key privacy and security layers with a 360-degree approach.



Privacy Preserving & Cyber Security Framework (HVD's)

The key processes constituting the privacy risk framework as inspired by NIST privacy framework are as follows:

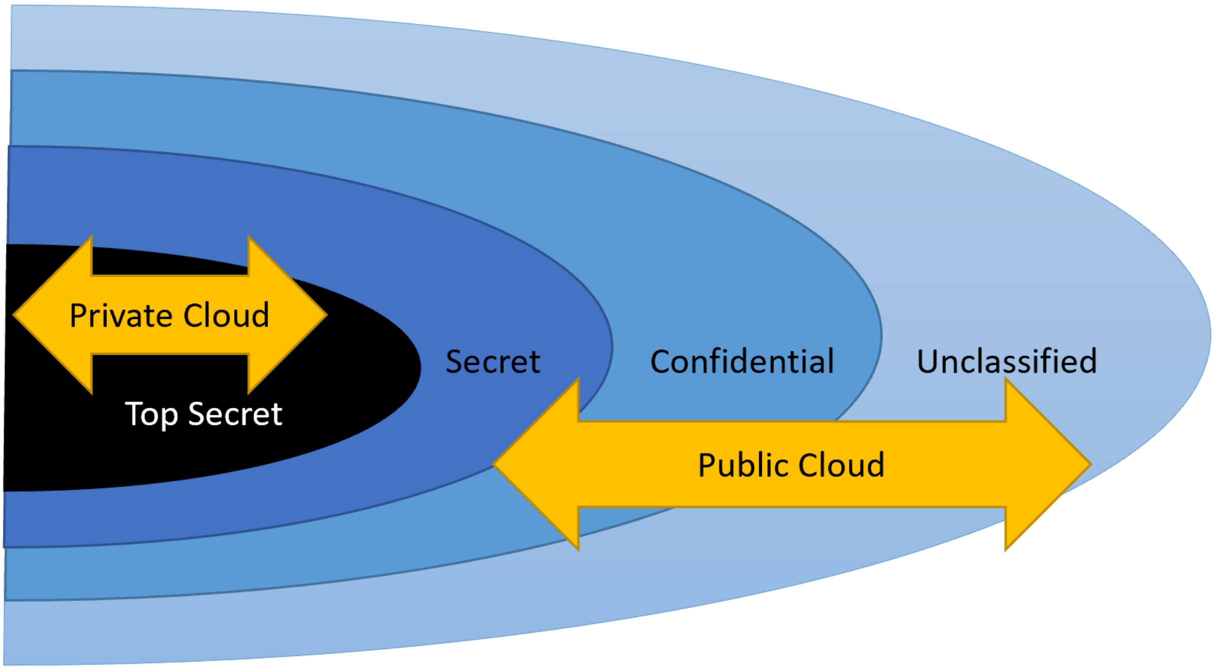
- Identify- Developing the understanding to manage privacy risk related to sharing HVD on the OGD platform. The activities in this process are foundational for effective use of the Privacy Framework. It includes inventorying the meta data related to the HVD along with classification and sensitivity labelling enabling the stakeholders to understand the risks involved and determine appropriate action to address the same.

- Govern- This process focuses on overall privacy values and policies, identifying legal/regulatory requirements as well as risk tolerance that enable CDOs to focus and prioritize their efforts in alignment to the defined risk management strategy and business needs.

- Control- This process involves developing and implementing specific activities to manage privacy risks including defining authorization policies pertaining to sensitivity labelling review, release/revoke of HVD on OGD platform, sharing methodology based on identified risk. The control process also involves activities leading to manageability of the HVD like detailing policies around deletion, modification, insertion of data attributes and audit logs related to usage. Finally, the process details the strategies that could be adopted to address the identified risk for example dissociated processing for data minimization (These are elaborated in the recommendation section).

- Communicate – Develop and implement appropriate activities to enable stakeholders to have a reliable understanding and engage in a dialogue about how HVD privacy risk management is conducted.

- Protect – This involves the entire set of activities directed at ensuring data protection to prevent cybersecurity-related privacy events, the overlap between privacy and cybersecurity risk management. The datasets can be preserved and secured based on sensitivity of data (confidential, secret, and top secret), and some of the tech options for sharing HVDs corresponding to each of these sensitivity labels are listed below and can be adopted based on the defined governance policies.



Tech Strategies to protect confidential HVDs

Key enabling technologies and services that customers may find helpful when working with confidential HVDs are listed below:

- All recommended technologies used for unclassified data, especially services such as Virtual Network, Intrusion detection Tools, Intrusion Prevention Tools, Log and Monitoring Tools.
- Allowing traffic through private connections over a private link so that data travels over a secure backbone eliminating the need to share sensitive HVD to the public internet.

- Data encryption at rest and in transit is recommended with customer-managed keys (CMK) backed by multi-tenant hardware security modules (HSMs) that have FIPS 140-2 Level 2 validation.
- For support scenarios adopt Just-in-Time (JIT) workflow that comes with full audit logging enabled.

Using these capabilities, customers can achieve the level of isolation, security, and confidence required to store confidential HVDs. Customers should use XDR Tools for threat detection and Monitoring capabilities for visibility into the usage and access patterns of HVDs including security posture.

Tech strategies to protect secret HVDs

Key enabling technologies and services that customers may find helpful when deploying secret data and workloads in public cloud are listed below:

- All recommended technologies used for confidential data.
- Use FIPS 140-2 Level 3 validated HSMs bound to a separate security domain controlled by the customer and isolated cryptographically from instances belonging to other customers.
- Physical servers that can host one or more VMs and are dedicated to one subscription. Customers can provision dedicated hosts, and can then place VMs directly into provisioned hosts using whatever configuration best meets their needs. Dedicated Host provides hardware isolation at the physical server level, allowing customers to place their VMs on an isolated and dedicated physical server that runs only their organization's workloads to meet corporate compliance requirements.
- Accelerated FPGA networking based on specialized NICs allows customers to offload host networking to dedicated hardware, enabling tunnelling for VNets, security, and load balancing. Offloading network traffic to a dedicated chip prevents side-channel attacks on the main CPU.
- Confidential computing offers encryption of data while it's in use, ensuring that data is always under customer control. Data is protected inside a hardware-based trusted execution environment (TEE, also known as enclave), and there is no way to view data or operations from outside the enclave.

To accommodate secret data in the public multi-tenant cloud, customers can deploy additional technologies and services on top of those used for confidential data and limit provisioned services to those that provide sufficient isolation. These services offer isolation options at run time and support data encryption at rest using customer-managed keys in dedicated single tenant HSMs that are solely under customer control.

Tech strategies to protect top secret HVDs

Key enabling products that customers may find helpful when deploying top secret HVDs are listed below:

- Use all recommended technologies for secret data.
- Deploy HVDs in physically isolated network for their highest classification data.
- Explore tactical edge deployments for limited or no connectivity, fully mobile requirements, harsh conditions requiring military specification solutions, and so on.
- User-provided hardware security modules (HSMs) allow customers to store their encryption keys and other secrets in HSMs deployed on-premises and controlled solely by customers.

Accommodating top secret HVDs will likely require a disconnected environment. Even though "air-gapped" networks do not necessarily increase security, many departments may be reluctant to store data with this classification in an internet connected environment. Today hyperscale cloud vendors offers an unmatched variety of public, private, and hybrid cloud deployment models to address each type of HVDs regarding the control of their data.

Model data sharing framework

As India cruises forward towards realizing its ambition of becoming $5 trillion economy, OGD platform is an import step towards creating an echo system for harnessing data fostering better governance, service delivery and innovation in sectors critical for inclusive social, economic, and sustainable growth. They objectives OGD platform are as follows:

Innovation

- Increasing the availability of high-value datasets of national importance
- Enabling secure pathways to share detailed datasets for research & development

Efficiency

- Improving policymaking, evaluation, and monitoring
- Enhancing the efficiency of service delivery
- Facilitating the creation of public digital platforms
- Streamlining inter-government data sharing

Openness

- Maximising access to and use of quality public sector data
- Promoting transparency, accountability, and ownership in data sharing & release
- Building digital & data capacity, knowledge & competency of government officials
- Ensuring greater citizen awareness, participation, and engagement with open data
- Increasing the availability of high-value datasets of national importance

Data quality

- Promoting data interoperability & integration to enhance data quality and usability
- Improving overall compliance to data sharing policies and standards

Security and privacy

- Protecting the privacy and security of all citizens
- Enabling secure pathways to share detailed datasets for research & development

It is imperative to formulate and share with CDOs a comprehensive data sharing framework for HVDs aligned with the core principles defined in India Data Accessibility and Use Policy released in Feb 2022:

Societal benefits

- Proactive data sharing for innovation & research
- Protection of intellectual property

One of the key focus areas while building the foundation of the data sharing framework is to look for the benefit of publishing data with purpose, by taking an active approach to sharing useable data. The HVD evaluation framework detailed earlier in the report provides guidelines and standards developed by the Open Government Data Platform to assess the value of datasets based on use cases, usage and market value. The framework is designed to help government departments and data providers identify high-value datasets, assess their impact, and improve their quality.

Security & accountability

- Risk management over risk avoidance
- Trust among stakeholders, systems, and transactions
- Privacy & security by design
- Well-defined accountability for all stakeholders
- Regulatory clarity & structured enforcement

A recurring theme from the case studies is the importance of fostering trustworthy data ecosystems and pursuing practical approaches for addressing privacy and cybersecurity challenges including applying risk-based data classification policies and adherence to domestic regulations. The privacy protection framework provides an integrated approach towards managing privacy and cyber security risks in alignment to regulatory compliances, international policies and practices defined in standard broadly accepted frameworks like NIST.

User centric & interoperable

- Interoperable, integrated & technology agnostic
- User-centred practices & systems
- Equal and non-discriminatory access

One of the fundamental objectives of this report is to drive increased adoption of the OGD platform. Several steps including an intuitive user interface, engagement with social media, implementing strategies that incentivize stakeholders to enter into data sharing efforts, ability to collaborate and engage via feedbacks and tools like AI driven automated support mechanism are envisaged to create a robust user centric ecosystem for harnessing the value of data.

Openness and transparency

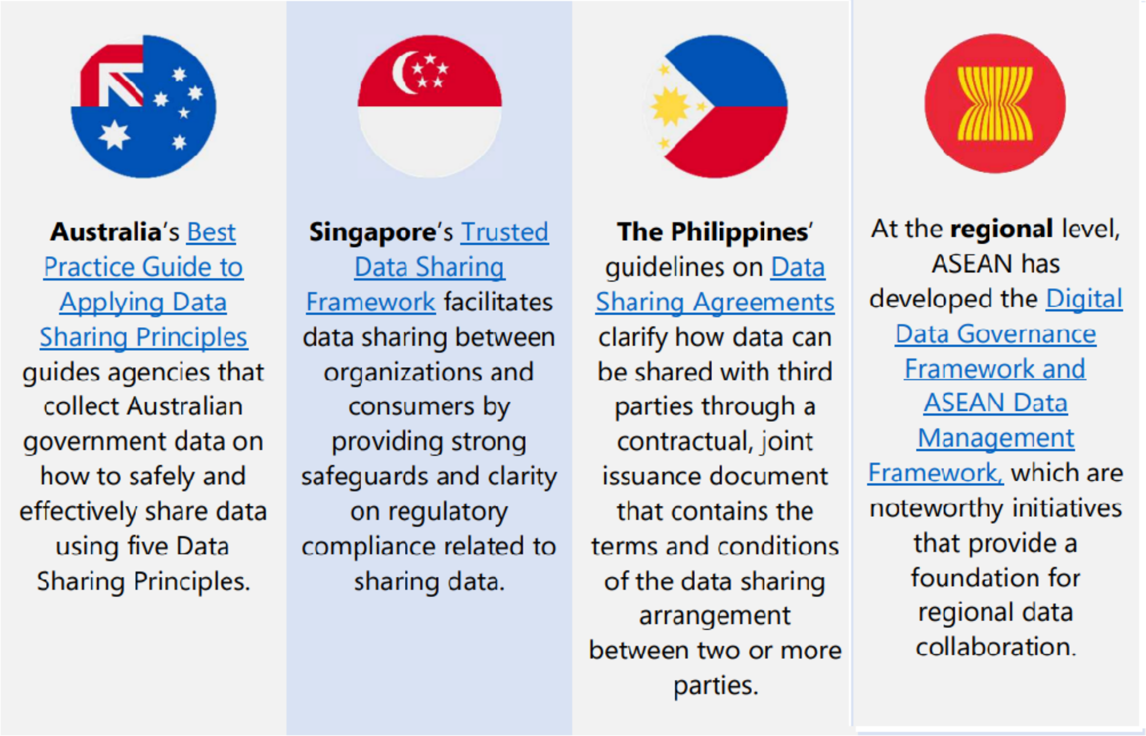
- Open by default
- Transparency in operations

The recommendation is to utilise tools and technologies which can facilitate government organizations to publish their datasets in open formats for free public use. This creates a valuable connecting link for government, citizens and the community to create an open data ecosystem in the country. Also, an increased impact of government data sharing efforts has been observed when there is collaboration with the private sector, the research community, non-government organizations and others. Availability of consistent, standardized data-sharing terms and licensing agreements enables seamless collaboration between such a large variety of stakeholders. In the context of OGD licensing and agreement for Open use of data in terms of agreement and licensing, there are three possible ways recommended

- Propose on open use of data agreement: Designed for use with open datasets which don't include personal data or data owned by a data provider. It is the most open and least restricted of the three first proposals.
- Computational Use of Data Agreement (C-UDA): Designed to define a use of data sets for AI training purposes which contain third-party materials. This is a contract for use with a database which includes open data but also some elements which are copyright-protectable (such as photos or snippets of text). It's for training an AI model but prohibits the republishing or redistributing of the protectable elements.
- Data use agreement for Open AI Model Development (DUA-OAI): Designed for underlying data with elements which may involve privacy or when data may be proprietary to the controller of the data.

The value that can be unlocked through quality datasets has led to growing momentum on data sharing in various countries. Accordingly, the various data sharing frameworks globally have been studied. One of the key benefits common in various frameworks is the focus on greater societal awareness of potential benefits of data sharing. How the collaboration by various entities can be successful in term of OGD implementation is given with an example as below:

In one of the case studies data sharing impacted by improving the policy making process of data value sets and usage, In the case study the demonstration supporting policymaking is LinkedIn’s Labour Market Insights. LinkedIn has shared insights with governments based on aggregated data from its 730 million members and 55 million companies. This has provided a granular and real-time analysis of both the supply and demand sides of the labour market. Governments are better positioned to implement more effective policies to address gaps between unfilled roles and workers looking for jobs, promote the development of in-demand skills in the labour market, and address skills-related inequities.



Australia's Best Practice Guide to Applying Data Sharing Principles guides agencies that collect Australian government data on how to safely and effectively share data using five Data Sharing Principles.

Singapore's Trusted Data Sharing Framework facilitates data sharing between organizations and consumers by providing strong safeguards and clarity on regulatory compliance related to sharing data.

The Philippines' guidelines on Data Sharing Agreements clarify how data can be shared with third parties through a contractual, joint issuance document that contains the terms and conditions of the data sharing arrangement between two or more parties.

At the **regional** level, ASEAN has developed the Digital Data Governance Framework and ASEAN Data Management Framework, which are noteworthy initiatives that provide a foundation for regional data collaboration.

In the context of India OGD there are various upcoming regulations like Data Protection Bill which can be referred to while establishing the data sharing framework. Although it may be useful to develop national standards and practices tailored to local contexts, aligning as much as possible to international best practices helps promote cross-border data sharing while enhancing the usage of data. For example, in the area of information security, the ISO27001 family of standards provides a clear global benchmark for security controls that address some of the security concerns related to data sharing should be adopted under OGD.

It is found that establishing a foundation of trust through privacy, security and governance can build a strong enabling environment, which is crucial to building momentum on data sharing - *It is proposed for OGD to build the foundation of the platform on "Data culture & data trust as its core principle".*

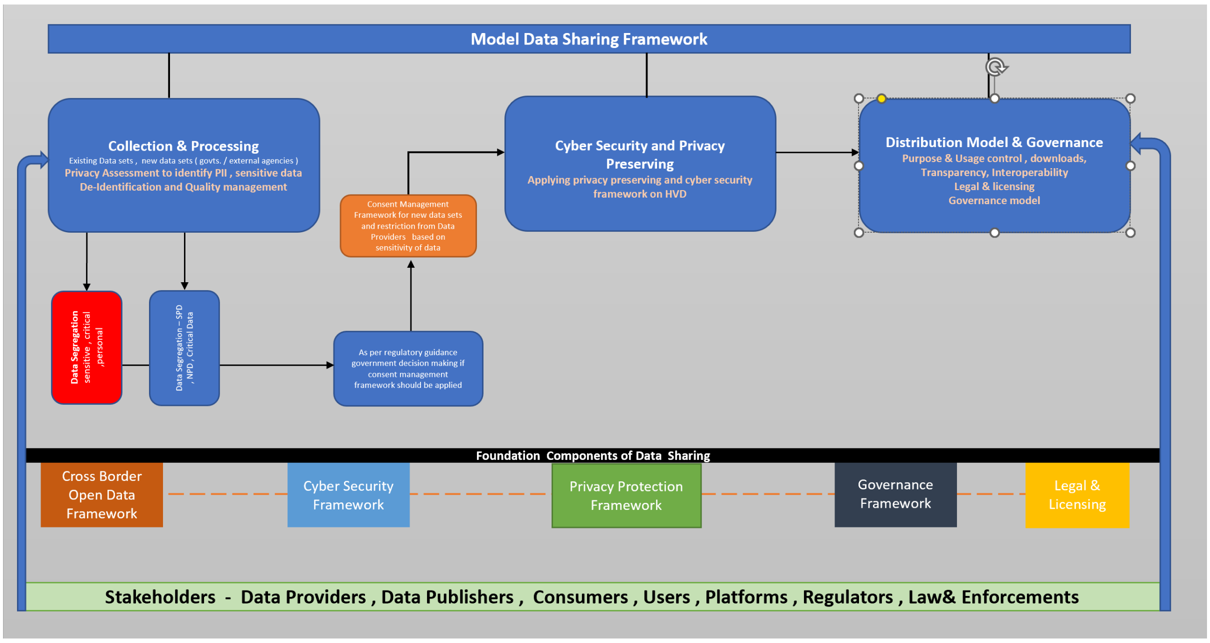
In order to further improve the data sharing model, three areas of impact were considered:

- **Economic:** The extent to which data sharing model stimulates innovation and competition, empowers data-oriented businesses to scale their operations and allows regional pooling of data to ensure businesses within the region remain internationally competitive.

- **Societal:** The extent to which data sharing model fosters trust and cooperation through improved transparency and accountability, deepens citizen engagement and encourages participation and responsibility, supports research and cooperation across communities to allow for joint discoveries.

- **Governance** - Governance in a model data sharing framework includes processes for improving quality of data, its interoperability, accuracy, and ensuring data sharing happens in an environment that ensures protection of data privacy, and security and regulatory compliance.

The diagram below depicts a data sharing model (high-level depiction) covering the flow for data processing and applying privacy preserving tools on the dataset based on their sensitivity classification.


Model Data Sharing Framework
Collection & Processing
Existing Data sets , new data sets (govts. / external agencies)
Privacy Assessment to identify PII , sensitive data
De-Identification and Quality management
Data Segregation
sensitive , critical ,personal
Data Segregation – SPD , NPD , Critical Data
As per regulatory guidance government decision making if consent management framework should be applied
Consent Management Framework for new data sets and restriction from Data Providers based on sensitivity of data
Cyber Security and Privacy Preserving
Applying privacy preserving and cyber security framework on HVD
Distribution Model & Governance
Purpose & Usage control , downloads,
Transparency, Interoperability
Legal & licensing
Governance model
Foundation Components of Data Sharing
Cross Border Open Data Framework
Cyber Security Framework
Privacy Protection Framework
Governance Framework
Legal & Licensing
Stakeholders - Data Providers , Data Publishers , Consumers , Users , Platforms , Regulators , Law& Enforcements

It is proposed to review if 'consent management based on regulatory aspects' are to be applied while collecting data from data providers, data owners, third party API's.

An important operational consideration of data sharing on OGD, where multiple parties contribute datasets for public good and for analytics to improve citizens services, is the framework structure of memorandum of understandings (MOUs), non-disclosure agreements (NDAs), codes of practice, intellectual property rights (IPRs), and contracts. These must conform to required data privacy and data security standards.

It is proposed to implement transparent, standardized, and risk-based data classification policies. This will ensure maximum sharing of open government data and that data is not unnecessarily prevented from being shared due to over-classification.

Key recommendations

Make open government data a policy priority, with a focus on sharing data in a way that is useable for data collaboration

Apply reasonable data classification policies to avoid the over-classification of data that could prevent the sharing of certain government data sets

Adoption of international standards and tools (such as ISO/IEC 38505 for data governance, ISO27701 for privacy information management, the NIST Cybersecurity Framework for cybersecurity, or the FAIR principles for scientific data management and stewardship)

Create regulatory sandboxes to allow for responsible testing of data sharing approaches in a “safe” space

A data sharing toolkit enabling automated discovery of existing datasets, their metadata including classification, a mechanism to get recommendations on policies to be applied basis sensitivity/classification and enforcement of the same will tremendously accelerate the pace at which CDOs can publish HVDs on the platform.

Proposed to provide non-binding guidance through regulators on how existing regulations apply in data sharing contexts

Recognize that while technical data-related skills are important, it is also important for policymakers and decision makers within organizations to have high level awareness of the opportunities that data sharing presents

2. Drive adoption of platform

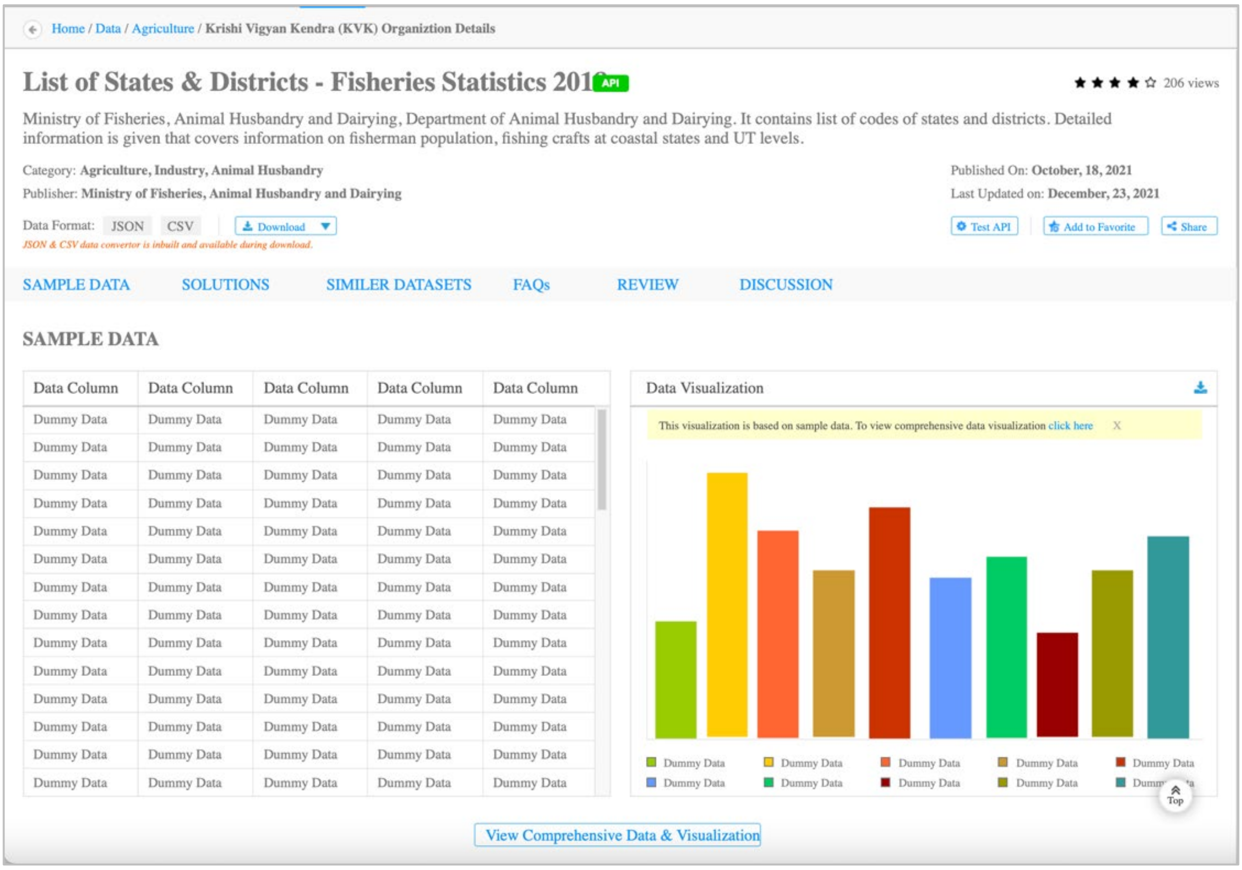
While improving quality and availability of datasets on the Open Government Data Portal is key to promoting the use of data and unlocking its true value, efforts need to be made to ensure that these datasets are hosted on a user-friendly and streamlined platform that is easy to navigate. To ensure this, the following recommendations are made to drive adoption of the OGD portal:
We recommend streamlining user navigation, reducing information overload, adding community engagement features, and establishing feedback loops to enhance overall user experience of the OGD platform in the report.

We recommend integration of the OGD platform with social media website, creating provision for testimonials and case studies for building trust and provision of chatbots, customer care & prominent FAQs for seamless user experience. We have shared detailed implementation roadmap of these recommendations along with wireframes suggesting redesign measures.

We designed wireframes based on the findings of our qualitative research, survey responses and heuristic evaluation. While MeitY has been working on their version 2.0 for the website, the following inputs are based on our research (as of Sept 2021- Jan 2022) that may be included by them in their new version.

Data visualization page

A detailed view of the data page, allows for preview and selection.



Home / Data / Agriculture / Krishi Vigyan Kendra (KVK) Organiztion Details

List of States & Districts - Fisheries Statistics 201 API

★★★★☆ 206 views

Ministry of Fisheries, Animal Husbandry and Dairying, Department of Animal Husbandry and Dairying. It contains list of codes of states and districts. Detailed information is given that covers information on fisherman population, fishing crafts at coastal states and UT levels.

Category: Agriculture, Industry, Animal Husbandry

Publisher: Ministry of Fisheries, Animal Husbandry and Dairying

Published On: October, 18, 2021

Last Updated on: December, 23, 2021

Data Format: JSON CSV

JSON & CSV data convertor is inbuilt and available during download.

SAMPLE DATA

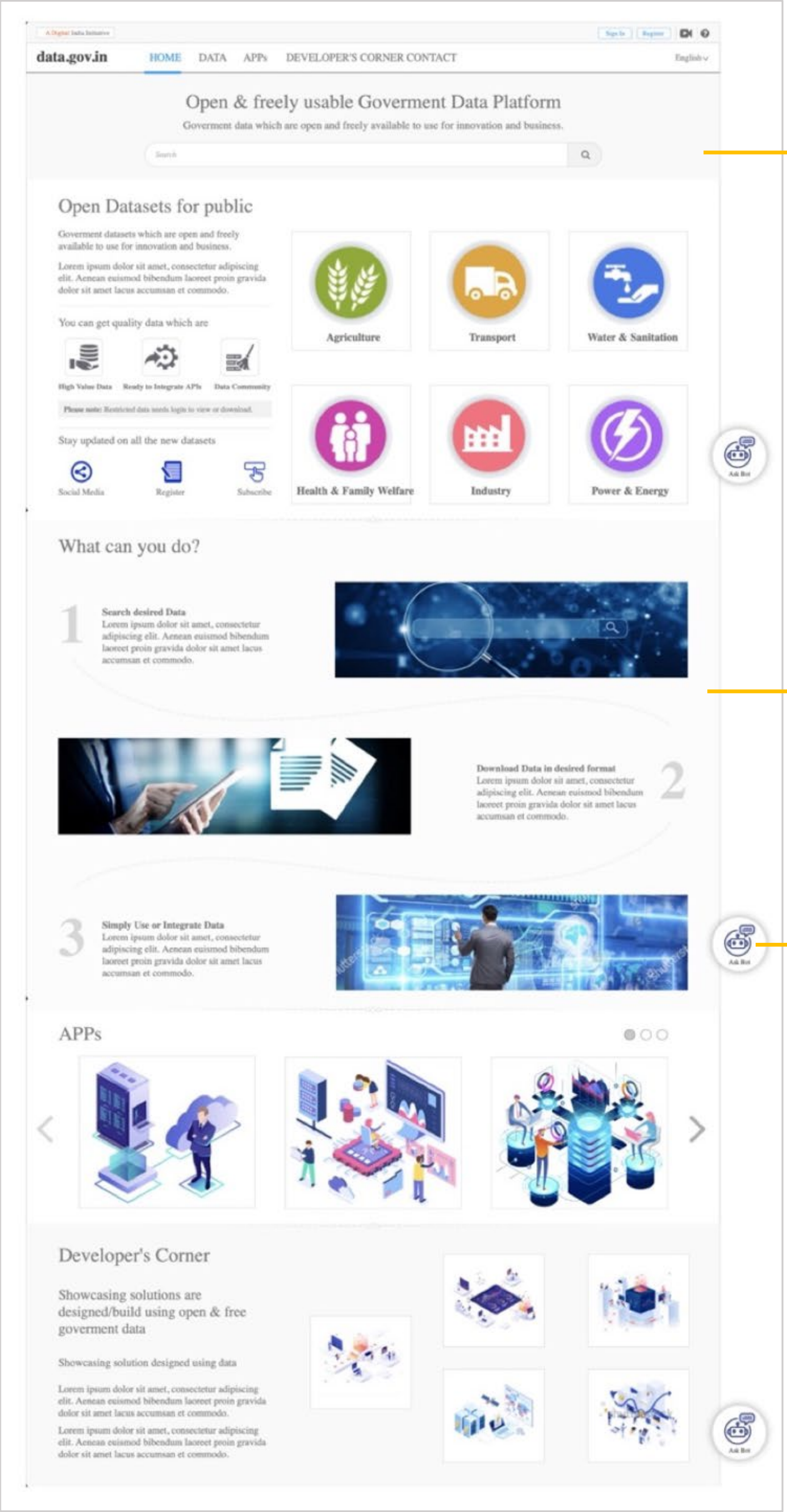
Data Column	Data Column	Data Column	Data Column	Data Column
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data

Data Visualization

This visualization is based on sample data. To view comprehensive data visualization click here

Home page

The homepage was redesigned to streamline the user's navigation across the platform ecosystem and to reduce the user's cognitive overload.



A Digital India Initiative

data.gov.in

Open & freely usable Goverment Data Platform

Goverment data which are open and freely available to use for innovation and business.

Open Datasets for public

Goverment datasets which are open and freely available to use for innovation and business.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet proin gravida dolor sit amet lacus accumsan et commodo.

You can get quality data which are

Please note: Restricted data needs login to view or download.

Stay updated on all the new datasets

What can you do?

1 **Search desired Data**
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet proin gravida dolor sit amet lacus accumsan et commodo.

Download Data in desired format 2
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet proin gravida dolor sit amet lacus accumsan et commodo.

3 **Simply Use or Integrate Data**
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet proin gravida dolor sit amet lacus accumsan et commodo.

APPs

Developer's Corner

Showcasing solutions are designed/build using open & free goverment data

Showcasing solution designed using data

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet proin gravida dolor sit amet lacus accumsan et commodo.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet proin gravida dolor sit amet lacus accumsan et commodo.

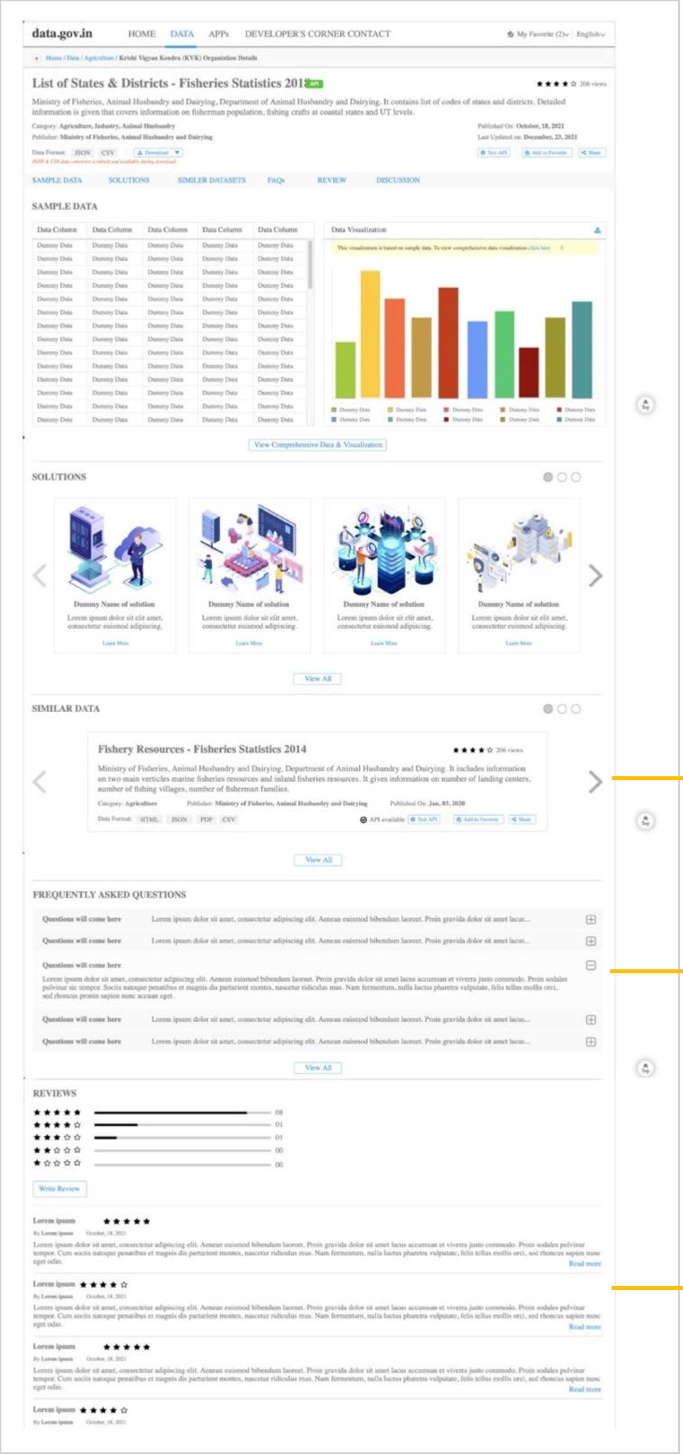
Streamlined user navigation

Reduced information overload

Chatbot feature

Category data

The proposed changes cover ease of use and accessibility, boosting meaningful engagement with the community, and establishing feedback loops to be demand driven. This interface allows users to explore all information about a specific catalog/dataset, including and not limited to community contributions and feedback, potential for reuse, details about collection methodology and metadata, etc.



Home / Data / Agriculture / Krishi Vigyan Kendra (KVK) Organization Details

List of States & Districts - Fisheries Statistics 2013 ★★★★☆ 206 views

Ministry of Fisheries, Animal Husbandry and Dairying, Department of Animal Husbandry and Dairying. It contains list of codes of states and districts. Detailed information is given that covers information on fisherman population, fishing crafts at coastal states and UT levels.

Category: **Agriculture, Industry, Animal Husbandry**
Publisher: **Ministry of Fisheries, Animal Husbandry and Dairying**
Published On: **October, 18, 2021**
Last Updated on: **December, 23, 2021**
Data Format: JSON CSV
JSON & CSV data conversion is inbuilt and available during download

SAMPLE DATA

Data Column	Data Column	Data Column	Data Column	Data Column
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data
Dummy Data	Dummy Data	Dummy Data	Dummy Data	Dummy Data

This visualization is based on sample data. To view comprehensive data visualization click here X

SOLUTIONS

Dummy Name of solution
Lorem ipsum dolor sit elit amet, consectetur euismod adipiscing.

Dummy Name of solution
Lorem ipsum dolor sit elit amet, consectetur euismod adipiscing.

Dummy Name of solution
Lorem ipsum dolor sit elit amet, consectetur euismod adipiscing.

Dummy Name of solution
Lorem ipsum dolor sit elit amet, consectetur euismod adipiscing.

SIMILAR DATA

Fishery Resources - Fisheries Statistics 2014 ★★★★☆ 206 views

Ministry of Fisheries, Animal Husbandry and Dairying, Department of Animal Husbandry and Dairying. It includes information on two main verticles marine fisheries resources and inland fisheries resources. It gives information on number of landing centers, number of fishing villages, number of fisherman families.

Category: **Agriculture** Publisher: **Ministry of Fisheries, Animal Husbandry and Dairying** Published On: **Jan, 03, 2020**
Data Format: HTML JSON PDF CSV API available

FREQUENTLY ASKED QUESTIONS

Questions will come here Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus... ⊞

Questions will come here Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus... ⊞

Questions will come here ⊟
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus accumsan et viverra justo commodo. Proin sodales pulvinar sic tempor. Sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nam fermentum, nulla luctus pharetra vulputate, felis tellus mollis orci, sed rhoncus pronin sapien nunc accuan eget.

Questions will come here Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus... ⊞

Questions will come here Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus... ⊞

REVIEWS

★★★★★	08
★★★★☆	01
★★★☆☆	01
★★☆☆☆	00
★☆☆☆☆	00

Write Review

Lorem ipsum ★★★★★
By **Lorem ipsum** October, 18, 2021
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus accumsan et viverra justo commodo. Proin sodales pulvinar tempor. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nam fermentum, nulla luctus pharetra vulputate, felis tellus mollis orci, sed rhoncus sapien nunc eget odio.
Read more

Lorem ipsum ★★★★☆
By **Lorem ipsum** October, 18, 2021
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus accumsan et viverra justo commodo. Proin sodales pulvinar tempor. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nam fermentum, nulla luctus pharetra vulputate, felis tellus mollis orci, sed rhoncus sapien nunc eget odio.
Read more

Lorem ipsum ★★★★★
By **Lorem ipsum** October, 18, 2021
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus accumsan et viverra justo commodo. Proin sodales pulvinar tempor. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nam fermentum, nulla luctus pharetra vulputate, felis tellus mollis orci, sed rhoncus sapien nunc eget odio.
Read more

Lorem ipsum ★★★★☆
By **Lorem ipsum** October, 18, 2021

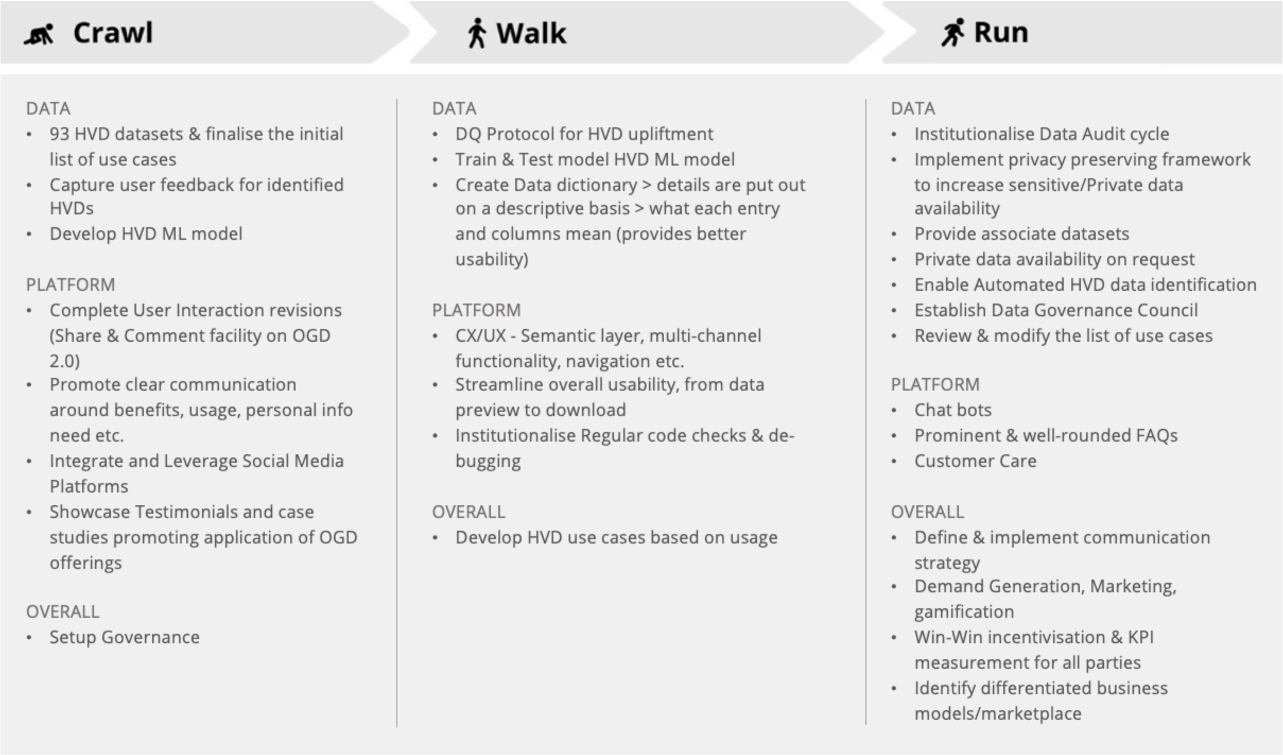
Provision for testimonials

Prominent FAQs

Community engagement feature

Roadmap

For structured and sustainable implementation of proposed solutions, the recommendations in this report, both incremental and strategic, have been streamlined into a prioritized roadmap for the Ministry of Electronics and Information Technology. We have streamlined the recommendation by data, platform & overall governance for relevant stakeholders:



Crawl

DATA
- 93 HVD datasets & finalise the initial list of use cases
- Capture user feedback for identified HVDs
- Develop HVD ML model

PLATFORM
- Complete User Interaction revisions (Share & Comment facility on OGD 2.0)
- Promote clear communication around benefits, usage, personal info need etc.
- Integrate and Leverage Social Media Platforms
- Showcase Testimonials and case studies promoting application of OGD offerings

OVERALL
- Setup Governance

Walk

DATA
- DQ Protocol for HVD upliftment
- Train & Test model HVD ML model
- Create Data dictionary > details are put out on a descriptive basis > what each entry and columns mean (provides better usability)

PLATFORM
- CX/UX - Semantic layer, multi-channel functionality, navigation etc.
- Streamline overall usability, from data preview to download
- Institutionalise Regular code checks & debugging

OVERALL
- Develop HVD use cases based on usage

Run

DATA
- Institutionalise Data Audit cycle
- Implement privacy preserving framework to increase sensitive/Private data availability
- Provide associate datasets
- Private data availability on request
- Enable Automated HVD data identification
- Establish Data Governance Council
- Review & modify the list of use cases

PLATFORM
- Chat bots
- Prominent & well-rounded FAQs
- Customer Care

OVERALL
- Define & implement communication strategy
- Demand Generation, Marketing, gamification
- Win-Win incentivisation & KPI measurement for all parties
- Identify differentiated business models/marketplace

This roadmap may be considered by MeitY as a step-by-step guide to improve India’s ability to harness the potential of data and to leverage it as a key social and economic resource with the potential of driving inclusive socio-economic progress in the country.

06 Appendix

1. OGD-user research methodology
2. Fractal NASSCOM open government data - insights-opportunities
3. Data sharing proposals comparison
4. Dataset audit findings
5. HVD identification framework
6. OGD feature roadmap
7. Privacy preserving & data sharing track

Detailed list: Top 93 datasets within various sectors, to be identified, uploaded & curated on the platform

Sector	Datasets	Description
Agriculture	Crop cycle plans	Information about crop cycle for different crop varieties and their yield, cost
	Markets	Agriculture markets linked to GIS codes
	Land ownership	Digital land records
	Irrigation	Irrigation data for better water and crop management
	Weather & climate	Up to date climate data as per location
	Crop Diseases & infestations	Information about crop diseases and guide to manage them
	Soil & minerals	Data of soil minerals as per region
Health	Infrastructure	Data of health infrastructure and available facilities
	Covid management and monitoring	Covid facility related data at one place
	Smoking, drugs & alcohol consumption	Data related to alcohol consumption and related diseases
	Medicines	Data on availability and use of different medicines
	Doctor specialization	Information about doctors and their specialisation
	Covid vaccination data	Data of covid vaccination and response to it
	Mortality	Mortality data with causes
	Cancer	Information cancer treatment facilities and mortality
Business & economy	Raw material utilization by sector	Details on raw material use by different business sectors
	Currency	Performance of different currencies over the time
	Power and water supply	Details on power and water supply for businesses
	Fuel prices	Fuel prices and changes over the time and region
	GST & taxation	tax data and changes within it
	Import & Export	Import export data of different commodities
	Manufacturing & trade	Data on production capacity of different manufacturing businesses
	Msmes	Information on facilities available to MSMEs
	Urban/rural lending demand	Credit pattern in urban and rural areas
Urban planning	Power supply	Data on power supply and consumption in urban areas
	Water supply	Data on water supply and consumption in urban areas
	Traffic data	GIS based traffic data for better traffic management
	Waste management	Information on waste management facilities

	Flood data	Flood data of different cities
	Roadways	Information on road connectivity in different areas
	Human-animal conflict	Data on incidents of human animal conflict in urban area
	Connectivity	Data on different connectivity modes in cities
	Land type	Different land types in urban region
Education & Research	Number & type of institutes	Details on education institutes in area
	Number of Students	Data on number of students in different education institutes
	Financial aides	Data on options of financial assistance to student for education
	University enrollments	Enrollment data in different universities
	Dropouts	Students drop out data for different education facilities
Insurance	# Of policies issued	Details of different insurance policies
	Public/private issuance	
	Average policy cost	Information on insurance policy costs
	Frauds	Information on insurance policy frauds
	Waste and abuse	Data on misuse of insurance services
	Reach	Information on level of reach of different insurance policies
	Industry analysis	Insurance analysis for different industry
Share market	Trading volume per market category	Data on trading trends of different securities
	# of listed companies	Data of listed companies and their details
	Foreign direct investments	FDA details as per different sectors
	Corporate bond issuances	Information on corporate bonds of different companies
	Market analysis	Analysis of market as per the demand and past trends
	Market participants	Trade data including exports and imports, tariffs based on market categories
Sports	Upcoming tournaments	Tournaments details like schedules, squads, teams etc.
	Sport records	Information on all sports events, categories, win/loss, fitness stats
	Youth enrollment	Eligibility criteria, age limitations for different sport categories enrollment for youth
	Recreational land availability	Land records available by states/districts for recreational activities
	Sports infrastructure	Stadiums, coaches, kits, training equipment availability and numbers of states/district wise
Environment & sustainability	Real time pollution data	Hourly air quality index data for each district
	Local weather data	Details of weather conditions - Temperature changes, wind speed, precipitation forecast
	Energy sources & consumption	Energy consumption units, Sources by districts, availability by categories
	Deforestation & green cover	Number of trees, green belts and other activist's details
	Wildlife	Species counts by states/districts, extinct numbers
	Natural disasters	Information on the past natural and man-made disasters, loss of life counts, frequency
	Policies, aides & funds	Funding from government by state and frequency for environmental impact
Technology	Cost	Money spent on technology adaptation by the government
	Usage	Data consumptions units, usage stats and state/district wise usage data
	Digital take-up	Data related to digitization using drones and robots available

	Completion rate	Information on installation of fiber and status of each area
	Satisfaction	Data for voice call quality, mobile data speed and internet quality
	Cyber safety	Data for past cyber-attacks, safety practices and monitoring infrastructure availability
Crime & justice	Infrastructure	Information on number of jails, police stations, staff, and equipment like PCRs
	Criminal records	Data related to crimes by state/district, digital records, and archives
	Crime rates/state-city	Information on crimes rate by types/category and frequency
	Communities related to prison reform	Information on communities by state/district for prison reform and inside prison conditions data
	Courts	Number of courts by state/district, judges, and cases historical and ongoing
	Crime against women & children	Data on past crimes against women and children, their frequency, and types
	Cybercrime	Information on past cyber-attacks, frequency, types, and security measures to mitigate
	Court Cases	Number of court cases, ongoing/pending/closed by state/district
	Laws	Information on all the criminal laws in place state/center, data on councils responsible for them
Transport	Airports	Data on number of planes, capacity, runway area, staff, and other security related infrastructure
	Roads	Data related to number of highways/expressways, road lengths, tolls details with pricing
	Freight	Number of goods transported by truck, train, planes, and taxes paid
	Parking	Parking slots available, area per sq foot, pricing for different vehicles
	Last mile connectivity	Modes of transport available to transfer to mainline like railway or bus from remote areas
	Modes available	Data on public and private modes of transport available, timings, vehicle type and frequency
	Railways	Data on number of trains, timings, connecting areas, distance, and ticket prices
	Public transportation	Information of public transport availability, types, duration, and prices
	Accidents	Data on frequency of accidents happening on which routes and by transport types
Government	Staff numbers and pay	Data on all the ministers in center/state government, their qualification, and demographic details
	Local counsellors and department	Data on number of counsellors department wise and their demographic details
	Business	Data on government businesses, ownerships, workforce, income and taxes from those sectors
	Policy making	All the policy documents related to government yojnas, five year plans, 3 year agendas
	Tenders	Information on government tenders schedule and date of submissions on tender portals
	Laws	Information on all the Laws in place state/union and data on councils responsible for them

Advisors:

Abhishek Singh, Advisor NeGD; MD Digital India Corporation

Deepika Raman, Project Lead; IIC

Khushal Wadhawan, Project Associate; IIC

Tushar Goel, Project Associate; IIC

Taskforce:

NASSCOM

Sangeeta Gupta, Sr. VP

Asna Siddiqui, Head, National AI Portal

Fractal

Srikanth Velamakanni, CEO

Sagar Shah, Client Partner

Avani Patel, Senior Design Consultant

Arunima Singh, Senior Design Consultant

Jay Amin, Lead Strategy Manager

Akbar Mohammed, Lead Data Scientist

Shagun Parab, Communication Designer

Dishali Patil, Consultant

Ashna Taneja, Consultant

Krutika Choudhary, Consultant

Microsoft

Dr. Rohini Srivathsa National Technology Officer

Deepak Talwar National Security Officer

Shikha Agrawal Chief Technology Officer – Data

Infosys

Gaurav Bhandari, AVP - Senior Principal

Saurabh Agarwal, Principal Consultant

Srinivas Amara, Senior Consultant

Saloni Grover, Consultant

Artha Global

Rajeswari Parasa Senior Analyst

Sridhar Ganapathy Senior Associate

TCS

Santanu Ghosh, Business Head

Sandeep Saxsena, Data Marketplace, Research Program Head

Tamanna Desai, UX Head

Joydeep Samajder, Enterprise Architect

Neha Sharma, Consultant

Santanu M, Senior Consultant

Amazon

Sainath Bandhakavi, Senior Solutions Architect

Sanjiv Jha, Principal Solution Architect