Draft Data Policy for Dharamshala Smart City

Ver. 1.0 (Draft)

June 2019

Dharamshala Smart City Limited
Old HIMUDA Building, First Floor,
In front of Rain Shelter, Chilgadi,
Dharamshala, Distt. Kangra - 176057

Table Contents

[redacted]

DHARAMSHALA SMART CITY LIMITED
OLD HIMUDA BUILDING, CHILGARHI, DHARAMSHALA, DISTRICT KANGRA, HIMACHAL PRADESH-176215.
Ph: 01892-226677, 226075 email – dscldharamshala@gmail.com, ceodsl@gmail.com

No. DSCL/Chief Data Officer/86/2018- 515-518 Dated: 15-07-2019

OFFICE ORDER

As per guidelines under Smart Cities Data Strategy and Data Maturity Assessment Framework the **Draft of Data Policy** in respect of The Dharamshala Smart City Limited approved by the MD-cum-CEO, DSCL Dharamshala is hereby released for uploading in the online portal of MoHUA as per deadline today.

(J. M. Awasthi)
Chief Data Officer & CFO,
Dharamshala Smart City Limited,
Dharamshala, Distt Kangra HP

15/7/19
Asstt Rec
Baldev

Copy to:

1. Sh. Kunal Kumar (IAS), Joint Secretary & Mission Director, Smart Cities Mission of MoHUA, New Delhi for information please.
2. MD-cum-CEO, DSCL Dharamshala for information please.
3. Data Officer & Accounts Officer, DSCL Dharamshala for information and n/a please.
4. Team Leader, PMC, DSCL, Dharamshala for information and n/a please.

(J. M. Awasthi)
Chief Data Officer & CFO,
Dharamshala Smart City Limited,
Dharamshala, Distt Kangra HP

1. History and Chronology of Data Policies in India

1.1 National Data Sharing and Accessibility Policy (NDSAP) – Government of India

The union cabinet approved the National Data Sharing and Accessibility Policy (NDSAP) on 9 February 2012. The objective of the policy was to facilitate access to Government of India (GoI) owned shareable data and information in both human readable and machine readable formats.

1.2 Introduction – Scope & Need for NDSAP

A large quantum of data generated by various organizations and institutions in the country remains inaccessible to the public, although most of such data is non-sensitive in nature and could be used by public for educational, scientific, socio-economic and developmental purposes. There has been an increasing demand by the community, that such data collected with the deployment of public funds should be made more readily available to all, for enabling rational debate, better decision making and use in meeting civil society needs. The NDSAP policy is designed to promote data sharing and enable access to Government of India owned data for national planning, development and awareness.

The National Data Sharing and Accessibility Policy (NDSAP) is applicable to all shareable non-sensitive data available either in digital or analogue forms but generated using public funds by various ministries, departments, subordinate offices, organizations, and agencies of Government of India as well as of the states. The objective of this policy is to facilitate access to Government of India owned shareable data through a wide area network, thereby permitting a wider accessibility and usage by public. The principles

on which data sharing and accessibility need to be based include: openness, flexibility, transparency, quality, security and efficiency.

Some state governments such as Sikkim, Madhya Pradesh, Telangana, Odisha and Assam have created their own state data policy on the lines of NDSAP.

Asset and value potentials of data are widely recognized at all levels. Data collected or developed through public investments, when made publicly available and maintained over time, their potential value could be more fully realized. There has been an increasing demand by the community, that such updated data collected with the deployment of public funds should be made more readily available to all, for enabling rational debate, increase transparency better decision making and use in meeting civil society and government needs.

Efficient sharing of data among data owners and inter-and-intra governmental agencies along with data standards and interoperable systems is the need of the hour. Hence, there was a need to formulate a policy on National Data Sharing and Accessibility which could provide an enabling provision and platform for proactive and open access to the data generated through public funds available with various ministries/departments/organizations of Government of India.

1.3 NDSAP: An Overview

The National Data Sharing and Accessibility Policy (NDSAP) is designed so as to apply to all sharable non-sensitive data available either in digital or analog forms but generated using public funds by various Ministries/Departments /Subordinate offices/Organizations/ Agencies of Government of India as well as States. The NDSAP policy is designed to promote data sharing and enable

access to Government of India owned data for national planning, development and awareness.

NDSAP aims to provide an enabling provision and platform for proactive and open access to the data generated by various Government of India entities. The objective of this policy is to facilitate access to Government of India owned shareable data (along with its usage information) in machine readable form through a wide area network all over the country in a periodically updatable manner, within the framework of various related policies, acts and rules of Government of India, thereby permitting a wider accessibility and usage by public.

2. Objectives

The objective of this policy is to facilitate the access to Government of India owned shareable data and information in both human readable and machine readable forms through a network all over the country in a proactive and periodically updatable manner, within the framework of various related policies.

3. Open Government Data Platform (OGDP)

The open government data initiative started in India with the notification of the National Data Sharing and Accessibility Policy (NDSAP), submitted to the Union Cabinet by the Department of Science and Technology, on 17 March 2012. The NDSAP identified the Department of Electronics & Information Technology as the nodal department for the implementation of the policy through National Informatics Centre (NIC), while the Department of Science and Technology continues to be the nodal department on policy matters. In pursuance of the Policy, the Open Government Data Platform India was launched in 2012.

OGD Platform India has been set up at https://data.gov.in to provide collated access to Resources (datasets/apps) under Catalogs, published by different government entities in open format. It also provides a search & discovery mechanism for instant access to desired datasets. OGD Platform also has a rich mechanism for citizen engagement. Besides enabling citizens to express their need for specific resource (datasets or apps) or API, it also allows them to rate the quality of datasets; seek clarification or information from respective Chief Data Officer.

OGD Platform has a strong backend data management system which can be used by government departments to publish their datasets through a predefined workflow. They can also create and/or view the visualizations created for their datasets. They shall also have a dashboard to see the current status on their datasets, visualizations, usage analytics as well as feedback and queries from citizens at one point.

https://data.gov.in domain is in existence since 2012, the site gone through a sea change in terms of bug fixes during its Alpha release, functionality enhancements after incorporating the suggestions received through an online Citizen Survey conducted between 5th November 2013 and 16th April 2014.

The current version launched on 11th December, 2014 is the stable release of the platform. After the launch of the Digital India programme in 2015, Open Government Data Platform India has been included as one of the important initiatives under "Information for All" – Pillar 6 (six).

OGD Platform also has a Communities component which facilitates forming of communities around datasets, domain of interest such as agriculture, education, health, or it could be app developers' community or even of data journalists. This shall give first hand input to development community for building new components, apps. It shall also give input to departments as

what kind of datasets is more useful and accordingly prioritize the release of the datasets.

"OGD Platform India" is a platform for supporting open data initiative of Government of India. The platform has been set up to provide collated access to resources (datasets/apps) under catalogs, published by different government entities in open format. The portal is intended to be used by Government of India ministries, departments and their organizations to publish datasets, documents, services, tools and applications collected by them for public use. The base "Open Government Data Platform India" is a joint initiative of Government of India and Federal government of the United States.

4. Government Open Data License

While appropriate open formats and related aspects for implementation of the policy has been defined in the "NDSAP implementation guidelines" prepared by an inter-ministerial task force constituted by the National Informatics Centre, the open license for data sets published under NDSAP and through the OGD Platform remained unspecified. The license Government Open Data License – India was published as an "Extraordinary Gazette" on February 2017. The license ensures that the data released are not misused or misinterpreted, and that all users have the same and permanent right to use the data.

Government Open Data License – India (GODL-India) is applicable to "all shareable non-sensitive data available either in digital or analog forms but generated using public funds by various agencies of the Government of India". The license permits anyone to "use, adapt, publish (either in original, or in adapted and/or derivative forms), translate, display, add value, and create derivative works (including products and services), for all lawful commercial and non-commercial purposes". In return, the user

must acknowledge the provider, source, and license of data by explicitly publishing the attribution statement, including the DOI (Digital object identifier), or the URL (Uniform Resource Locator), or the URI (Uniform Resource Identifier) of the data concerned.

1. The license does not cover the following kinds of data: Personal information
2. Data that is non-shareable and/or sensitive
3. Names, crests, logos and other official symbols of the data provider(s)
4. Data subject to other intellectual property rights, including patents, trade-marks and official marks
5. Military insignia
6. Identity documents
7. Any data that should not have been publicly disclosed for the grounds provided under section 8 of the Right to Information Act, 2005.

The key features are detailed as below:

1. **Open Source Driven** – Developed completely using Open Source Stack, facilitating cost saving in terms of software and licenses and also provisioning community participation in terms of further development of product with modules of data visualization, consumption, APIs to access datasets, etc.
2. **Metadata** – Resources (Datasets/Apps) shall be published along with standard metadata along with controlled vocabularies on government sectors, jurisdictions, dataset types, access mode etc. Besides facilitating easy access to datasets, this shall be extremely useful in the future for federation/integration of data catalogs.
3. **Social Media Connect**–To support wider reach and dissemination of datasets, anyone can share the information about any dataset published on the platform with his/her social media pages on a press of a click.

4. **Citizen Engagement** – The Platform has also a strong component of Citizen Engagement. Citizens can express their views as well as rate the datasets w.r.t three aspects (Quality, Accessibility and Usability) on the scale of 5. They can also embed the Resources (Datasets/Apps) in their blogs or web sites. Facility to contact the Chief Data Officer is also available on the Platform.
5. **Community Collaboration** – Citizens with specific interest can build communities and discuss online. OGD Platform facilitates the communities to open up online forums, blogs and discussions around various datasets, apps available on the platform. It also provides a platform to express and discuss the kind of Datasets, APPs & APIs they would like to have. It shall also give input to departments as what kind of datasets is more useful and accordingly prioritize the release of the datasets.
6. **Government Open Data Licence-India** – In the Task Force meeting held on September 10, 2015, the following decisions on Commercial use of Open Data has been taken: i) A committee to be formed under the Chairmanship of JS (Legal) & Chief Data Officer, D/o Legal Affairs with members from concerned departments to review and recommend the license to be associated with the data being published on the OGD Platform as well as its use for commercial purposes.

4.1 Technological aspects and guidelines for National Data Policy

NDSAP recommends that datasets has to be published in an open format. It should be machine readable. Considering the current analysis of data formats prevalent in Government, it is proposed that data should be published in any of the following formats:

1) CSV (Comma separated values)

2) XLS (Spread sheet - Excel)
3) ODS/OTS (Open Document Formats for Spreadsheets)
4) XML (Extensive Markup Language)
5) RDF (Resources Description Framework)
6) KML (Keyhole Markup Language used for Maps)
7) GML (Geography Markup Language)
8) RSS/ATOM (Fast changing data/feeds e.g. hourly/daily)

5. Classification of data

5.1 Different types of data

Different types of datasets generated both in geospatial and non-spatial form by different ministries/departments are supposed to be classified as shareable data and non-shareable data. Data management encompasses the systems and processes that ensure data integrity, data storage and security, including metadata, data security and access registers. The principles on which data sharing and accessibility need to be based include: Openness, Flexibility, Transparency, Quality, Security and Machine-readable.

1. Geospatial Data
2. Non-spatial Data

Datasets generated both in form by Ministries/Departments shall be classified as:

1. Shareable data and
2. Non-shareable data.

The derived statistics like national accounts statistics, indicators like price index, databases from census and surveys are the types of data produced by a statistical mechanism. However, the geospatial data consists primarily of

satellite data, maps, etc. In such a mechanism, it becomes important to maintain standards in respect of metadata, data layout and data access policy.

5.2 Types of Access

There are many types of access for datasets

1. **Open Access** – Access to data generated from public funding should be easy, timely, user-friendly and web-based without any process of registration / authorization.
2. **Registered Access –** Datasets which are accessible only through a prescribed process of registration / authorization by respective departments / organizations will be available to the recognized institutions / organizations / public users, through defined procedures.
3. **Restricted Access** – Data declared as restricted, by Government of India policies, will be accessible only through and under authorization.
4. **Technology for Sharing and Access -** State-of-the-art data warehouse and data archive with online analytical processing (OLAP) capabilities, which includes providing, a multi-dimensional and subject oriented view of the database needs to be created. This integrated repository of data portals of various ministries / departments as a part of https://data.gov.in will hold data and this repository over a period of time will also encompass data generated by various State Governments and UTs. The main features of the data warehouse need to include:

    1 User friendly interface
    2 Dynamic / pull down menus
    3 Search based Report
    4 Secured web access
    5 Bulletin board
    6 Complete Metadata
    7 Parametric and Dynamic report in exportable format

NDSAP once implemented would lead to:

1. Opening up of the information out of the Government System
2. Availability of Accurate, Reliable and Unbiased information to all
3. Providing single data OGD Platform for the country for data sharing
4. Establishment of a platform to promote innovation in government applications
5. Enhancing government Transparency, Accountability and Public Engagement
6. Effective utilization of Government data by providing meaningful visual representations.
7. Enabling development of Innovative Applications around datasets or mash-up from multiple datasets hence giving different perspectives to government data.

6. Metadata elements for catalog/resources

6.1 Data Source Formats for Metadata Discovery

Programming languages or Scripts or SQL queries can be used to write some programs to fetch the data from the relational database or some other sources to generate the desired metadata and dataset information. The https://data.gov.in tool can fetch the data catalog information with the help of the programming language. The language passes pre-defined metadata information to the data import tool to create the data catalogs automatically without any human interface. Datasets and related metadata may be created from a variety of different sources including:

1) Relational databases
2) CSV files or Spread sheets
3) XML files
4) Web Services or data in JSON format

6.1.1 Relational Databases

All most all the eGov applications use a relational database to store their reports and MIS data. Stored procedures or Database Triggers can be written to periodically export the data in to a dataset in CSV or XML format which can be consumed by https://data.gov.in data import tool to create the dataset catalog.

6.1.2 CSV Files or Spread sheets

CSV files or Spread sheets are the most efficient and common source of https://data.gov.in data import tool. eGov applications can specify the location from which data import tool has to fetch the dataset and metadata. Metadata and dataset can be put in a zip file with predefined naming convention. With the help of the tool https://data.gov.in can traverse designated location and extract the metadata and dataset in CSV files or spread sheets to convert it into data catalogs.

6.1.3 XML files

Structured XML files are one of the good sources for the data import. eGov applications can pass the metadata information and the details about the metadata in a predefined XML file. As defined by the eGov application https://data.gov.in data import tool can traverse the designated location of the XML at predefined time to fetch the data and convert it into searchable human readable data catalogs.

6.1.4 Web Services or Data in JSON Format

Web service can be used as a stream of data or services. https://data.gov.in supports both REST and SOAP based web service over HTTP. Using REST

based web services provides more flexibility to share data over XML as well as JSON format. eGov application has to submit the definition of the fields to be exposed by the web services which need to be mapped with the metadata information to create the catalog. At the given period of time with the help of metadata mapping the tool will import the data from the source apps and will convert to data catalogs.

6.2 Consumption of Datasets from e-Gov Applications

Metadata discovery is a way to fetch datasets and metadata information from different eGov applications. Desired eGov applications have to define the source and type of tool to be used to fetch dataset and metadata from their servers. The https://data.gov.in tool would crawl at source at specified time to fetch metadata/datasets.

6.3 Catalog

Title (Required): A unique name of the Dataset viz. Current Population Survey <Year>, Consumer Price Index <Year>, Variety-wise Daily Market Prices Data, State-wise Construction of Deep Tube wells over the years, etc.

Description (Required): Provide a detailed description of the Dataset e.g., an abstract determining the nature and purpose of the catalogs.

Keywords (Required): It is a list of terms, separated by commas, describing and indicating at the content of the dataset. Example: rainfall, weather, monthly statistics.

Group Name: This field allows agencies to provide a Group Name to closely related catalogs in order to show that they may be

presented as a group or a set.

Sector & Sub-Sector (Required): Choose the sectors(s)/sub-sector(s) those most closely apply (ies) to your catalogs.

Asset Jurisdiction (Required): This is a required field to identify the exact location or area to which the catalogs and resources (dataset/apps) caters to viz. entire country, state/province, district, city, etc.

6.4 Resources (Datasets/Apps)

Category (Required): Choose from the drop down options. Is it a Dataset or an Application?

Title (Required): A unique name of the resource viz. Consumer Price Index for etc.

Access Method (Required): This could be "Upload a Dataset" or "Single Click Link to Dataset".

Reference URLs: This may include description to the study design, instrumentation, implementation, limitations, and appropriate use of the dataset or tool. In the case of multiple documents or URLs, please delimit with commas or enter in separate lines.

If Resource Category is Dataset Frequency (Required): It mentions the time interval over which the dataset is published on the OGD Platform on a regular interval (one-time, annual, hourly, etc.).

Granularity of Data: It mentions the time interval over which the data inside the dataset is collected/ updated on a regular basis (one-time, annual, hourly, etc.).

Access Type: It mentions the type of access viz. Open, Priced, Registered Access or Restricted Access (G2G).

If Resource Category is App Type (Required): It mentions the type of App being contributed viz. Web App, Web Service, Mobile App, Web Map Service, RSS, APIs etc.

Data Sets Used: Datasets used for making this app.

Language: Language used for app.

Date Released: It mentions the release date of the Dataset/App.

Note: It mentions the anymore information the contributor/controller wishes to provide to the data consumer or about the resource.

NDSAP Policy Compliance: This field is to indicate if this dataset is in conformity with the National Data Sharing and Access Policy of the Govt. of India.

Best Practises for Catalogs/Resources/Datasets: Some of the best practices for contribution to be followed are stated as follows:

- Data should be stored in widely used file formats that are suitable for machine processing.
- Released dataset should clearly reflect "what is recorded about a particular subject".
- Timely release of datasets is one of the important factors to maximize the utility of information people can obtain.
- Data should be provided in freely available formats which can be accessed without the need for a software license.
- Data elements should be in de-normalized form

7. Guidelines, Procedures, Activities and Process for setting up NDSAP

In order to implement NDSAP, the Ministries/Departments of Government of India have to undertake the following activities:

- Nominate Chief Data Officer, at the level of Joint Secretary or above
- Chief Data Officer to nominate Data Contributors for coordinating from respective Divisions/Units/Projects/Schemes, etc.
- Create Data Contributor's login id. using Chief Data Officer's login account
- Setup NDSAP Cell
- Identify Datasets
- Prepare Negative List
- Publish Datasets on OGD Platform India
- Create Action Plan for regular release of datasets on the data.gov.in
- Monitor and Manage the Open Data Program of the Department

8. Nomination of Chief Data Officer

As per the directive from the Cabinet Secretary, a senior officer at the level of Joint Secretary or above is to be nominated as Chief Data Officer or Nodal Officer for the Department/Organization.

8.1 Role and responsibilities of Chief Data Officer

The main role & responsibilities of a Chief Data Officer are defined as follows:

- Head the NDSAP Cell, which helps in compilation, collation, conversion and publishing catalogs/resource on the platform. The size of the cell varies from Department to Department and depends on the quantum of resources to be published.
- Lead the open data initiative of Department/Organization
- Nominate Data Contributors

- Create Data Contributors login id using Chief Data Officer's login account.
- Take initiative to release as many datasets as possible on proactive basis.
- Identify the High Value Datasets and schedule their release on OGD Platform.
- Prepare the Negative List for the Department as per the directions in NDSAP.
- Ensures that the datasets being published, through a workflow process, are in compliance with NDSAP.
- Periodically monitor the release of datasets as per predefined schedule.
- Take relevant action on the feedback/suggestion received from the citizen for the datasets belonging to the Ministry/Department /Organization.
- Ensure the correctness of his contact details on the OGD Platform by sending a mail/ letter to ndsap@gov.in (email ID) in case of any change.
- Take action on Suggestions on new datasets made by public through the OGD Platform.

8.1.1 View & Respond to Queries on Published Datasets

Citizens can browse, search, filter, sort and access the datasets on the OGD Platform. Citizens also have the option to send their queries and feedbacks about the published datasets. This feedback would be available on the dash board of the Chief Data Officer to take further necessary action.

8.1.2 Respond to Suggestions for new Datasets

The OGD Platform has a strong Citizen Engagement feature built in. While browsing through the catalogue of datasets, if one is not able to find the dataset which is of interest to him then he can request for the same through suggestions module. Suggestions already made for particular datasets are displayed and one can also endorse the same. The suggested list i.e. the requirement for new datasets is sent to the respective department's Chief Data Officer. This would facilitate the Chief Data Officer to prioritize his release of datasets on the platform. They are expected to send in their response on the same.

8.1.3 Review Analytics & Plan

The Dash Board of the Chief Data Officer metrics would be available for the datasets contributed by all the contributors of that Ministry/Department/State. Feedback related to datasets would also be available along with the suggestions from citizen with respect to the requirement of new datasets. This feature would facilitate him to watch the analytics and accordingly plan his course of action.

8.1.4 DOs for Data Contribution and Approval

1) Identify and prioritize the release of datasets; categorize the type of access granted for them and publish as many high value datasets as possible.
2) Contribute datasets which are in the Open List and do not fall under the Negative List.
3) Ensure that the quality standards are met i.e. accuracy, free from any sort of legal issues, privacy of an individual is maintained and does not compromise with the National security.
4) Ensure that the datasets being published through a workflow process are in compliance with NDSAP. Details on original source of the dataset and

methodology of the data collection should be provided in metadata.

5) Prepare and contribute the metadata in predefined format for the Catalogs and Resources (Datasets/Apps). The key metadata elements are Title, Description, Category, Sector/Sub-Sector, Dataset Jurisdiction, Keywords, Access Method, Reference URLs, Data Group Name, Frequency, Granularity of Data and Policy Compliance. All the metadata elements must be filled with utmost quality and ease of use.
6) Pricing of data, if any, would be decided by the data owners as per the government policies.

9. NDSAP Cell

In order to implement NDSAP each Department would establish a NDSAP Cell. The size of the cell would vary from Department to Department and would depend on the quantum of datasets to be published. The NDSAP Cell would be responsible for:

1) Prepare Negative List of datasets and communicate to DST within Six Months
2) Prepare a schedule of datasets to be released in next one year
3) Extend Technical Support for Preparation of datasets, conversion of formats etc.
4) Monitor and manage the Open data initiative in their respective Ministry/ Department/State and ensure quality and correctness of the data
5) Work out an open data strategy to promote proactive dissemination of datasets
6) Institutionalize the creation of datasets as part of routine functioning

NDSAP Cell shall be headed by Chief Data Officer who could be assisted by number of Data Contributors. NDSAP Cell shall have professionals from data analyst, visualization and programming domain. The policy mentions that

budgetary provisions and appropriate support for data management for each department/organization would be necessary.

10. Data Contributor

In order to cater to the contribution of the datasets from offices/organization under the Ministries/Departments, the Chief Data Officer can nominate a number of Data Contributors who would be responsible in contributing the datasets along with their metadata. Using the web based DMS, each data contributor would be able to contribute the data as per the given metadata format which is based on the Dublin Core Standards. The contributed datasets would be approved by the Chief Data Officer as the case may be.

1) Data Contributor could be an officer of the Ministry/Department/State who would be responsible for his/her unit/division. The responsibilities of the Data Contributor are as follows:
2) Responsible for ensuring quality and correctness datasets of his/her unit/division
3) Preparing and contributing the catalogs and resources along with the metadata on the OGD Platform.

11. Publishing & Management of Resources (Datasets/Apps)

Contribution of datasets/apps is by login into a simple web based Dataset Management System through https://data.gov.in. Resources to be contributed under Catalogs are processed through a predefined workflow, ensuring compliance with government policies. Chief Data Officers nominated by government ministries or department are authorized to publish datasets in open format on OGD Platform.

Chief Data Officers of the Ministry/Department has the facility to create any number of Contributors for contributing Datasets/Apps for their Ministry/Department. Once the Contributor is created by the Chief Data Officer, a mail is sent to the mail id of the contributor. The Contributor then can login and contribute datasets along with its metadata for further approval by the Controller. However, the responsibility on the relevancy and quality of datasets published on the OGD Platform rests with Chief Data Officer.

12. Open Government Data (OGT)

"A dataset is said to be open if anyone is free to use, reuse, and redistribute it – Open Data shall be machine readable and it should also be easily accessible."

Government collects processes and generates a large amount of data in its day-to-day functioning. But a large quantum of government data remains inaccessible to citizens, civil society, although most of such data may be non-sensitive in nature and could be used by public for social, economic and developmental purposes.

These data need to be made available in an open format to facilitate use, reuse and redistribute; it should be free from any license or any other mechanism of control. Opening up of government data in open formats would enhance transparency and accountability while encouraging public engagement. The government data in open formats has a huge potential for innovation building various types of Apps, mash-ups and services around the published datasets.

Notification of NDSAP mandates government departments to proactively open up data. NDSAP in India is applicable to all entities of Government Setup.

13. Identification of Resources (Datasets/Apps) and their organization under Catalogs

As per policy each department has to prepare its Negative List. The datasets which are confidential in nature and are in the interest of the country's security in not opening to the public would fall into the negative list. This list would need to be compiled and sent to the Department of Science and Technology within six months.

All other datasets which do not fall under this negative list would be in the Open List. These datasets would need to be prioritized into high value datasets and non-high value datasets.

13.1 Data Categorization

There are four types of Data categorization:

1. Category 1: Data that may be freely disclosed with the public.
   Example: Contact information, price lists
2. Category 2: Internal data that is not meant for public disclosure.
   Example: Sales contest rules, organizational charts
3. Category 3: Sensitive internal data that if disclosed could negatively affect operations.
   Example: Contracts with third-party suppliers, employee reviews
4. Category 4: Highly sensitive corporate and customer data that if disclosed could put the organization at financial or legal risk.
   Example: Employee social security numbers, customer credit card numbers

As per the NDSAP, within a year all the datasets need to be published on the OGD Platform (https://data.gov.in) and within the first three months at least 5 high value.

The data which are contributed to the Open Government Data (OGD) Platform India have to be in the specified open data format only. The data have to be internally processed to ensure that the quality standard is met i.e. accuracy, free from any sort of legal issues, privacy of an individual is maintained and does not compromise with the National security. While prioritizing the release of datasets, one should try to publish as many high value datasets. Grouping of Related Resources (Datasets/Apps) should be planned and are to be organized under Catalogs.

Though each department shall have its own criterion of high value and low value datasets, generally High value data is governed by following Principles:

1. Completeness
2. Primary
3. Timeliness
4. Ease of Physical and Electronic Access
5. Machine readability
6. Non-discrimination
7. Use of Commonly Owned Standards
8. Licensing
9. Permanence
10. Usage Costs

14. Data Implementation Guidelines

(Tim Berners-Lee had classified data into single star to five star categories based on formats for the data. Please refer to Annexure II in the Implementation Guidelines)

1) Ensure that the data being uploaded on the OGD Platform is as complete as possible, reflecting the entirety of what is recorded about a particular

subject and is de-normalized. The datasets also should be optimized by adding redundant data or by grouping data before uploading.

2) Priority should be given to data whose utility is time sensitive. Real time information updates would maximize the utility the public can obtain from this information.
3) Replace any Not Available, Not Reported or missing values in the data with 'NA'.
4) Metadata that defines and explains the raw data should be included as well, along with formulas and explanations for how derived data was calculated.
5) Keywords must be defined in data catalog to make it search friendly.
6) Provide link to the reference documents (if any) or website for detailed information and explanation on the method of calculation or source of data.
7) Read the process manual at https://https://data.gov.in/help for the step by step procedure for contribution.
8) Prioritize the release of datasets and take relevant action on the basis of feedbacks and suggestions received on the OGD Platform from citizen's pertaining to the Ministry/Department.
9) Ensure the correctness of login details on the OGD Platform by sending a mail to ndsap@gov.in, in case of any change.

15. DON'Ts for Data Contribution and Approval

1) Don't contribute datasets which fall under the negative List e.g. the datasets which are confidential in nature and are in the interest of the country's security.
2) Don't impose 'Terms of Service', attribution requirements, restrictions on dissemination and so on, which act as barriers to public use of data.

3) Don't impose cost on the public for access of datasets, as imposing fees for access skews the pool of who is willing (or able) to access information.
4) Don't publish hand written note, as it is very difficult for machines to process. Scanning text via Optical Character Recognition (OCR) results in many matching and formatting errors. Information shared in the widely used PDF format is very difficult for machines to parse. Hence, the data in these formats should be avoided.
5) Data in non-Unicode formats should be avoided.
6) Don't contribute datasets with any special characters (e.g. @, %, $, &, etc.) or missing values.
7) Don't provide any explanation, including the method of calculation or source of data in data file to be attached in the web form.

16. Capacity Building

Two types of training modules both as offsite and onsite models have been envisaged. Each module would be for the duration of 2-3 days. The logistics and venue for the onsite training would be the responsibility of the host organization. The modules would be:

1) Awareness and Sensitization Module – for Chief Data Officer & other senior officers of the Ministries/Departments
2) Data Contribution Module – hands-on training for contributing datasets to the OGD Platform, provide advisory on conversion of data to digital format to Data Contributors and Members of NDSAP Cell.

17. National Data Sharing and Accessibility Policy (NDSAP) – 2012

17.1 Preamble

1.1 Asset and value potentials of data are widely recognized at all levels. Data collected or developed through public investments, when made publicly available and maintained over time, their potential value could be more fully realized. There has been an increasing demand by the community, that such data collected with the deployment of public funds should be made more readily available to all, for enabling rational debate, better decision making and use in meeting civil society needs. Principle 10 of the United Nations Declaration on Environment and Development (Rio de janeiro, June 1992), stated

"…….each individual shall have appropriate access to information concerning the environment that is held by public authorities …….

And

The opportunity to participate in the decision-making process. States shall facilitate and encourage public awareness and participation by making information widely available."

Section 4(2) of the Right to Information Act, 2005 reads

"It shall be a constant endeavor of every public authority to take steps in accordance with the requirements of clause (b) of sub-section (1) to provide as much information suomotu to the public at regular intervals through various means of communication, including internet, so that the public have minimum resort to the use of this Act to obtain information"

1.2 The principles on which data sharing and accessibility need to be based include: Openness, Flexibility, Transparency, Legal Conformity, Protection of Intellectual Property, Formal

Responsibility, Professionalism, Standards, Interoperability, Quality, Security, Efficiency, Accountability, Sustainability and Privacy.

1.3 A large quantum of data generated using public funds by various organizations and institutions in the country remains inaccessible to civil society, although most of such data may be non-sensitive in nature and could be used by public for scientific, economic and developmental purposes. The National Data Sharing and Accessibility Policy (NDSAP) is designed so as to apply to all sharable non-sensitive data available either in digital or analog forms but generated using public funds by various Ministries / Departments /Subordinate offices / organizations / agencies of Government of India. The NDSAP policy is designed to promote data sharing and enable access to Government of India owned data for national planning and development.

17.2 Definitions

2.1 Data –Data means a representation of information, numerical compilations and observations, documents, facts, maps, images, charts, tables and figures, concepts in digital and/or analog form.

2.2 Data Archive – A place where machine-readable data are acquired, manipulated, documented, and distributed to others for further analysis and consumption.

2.3 Data Generation– Initial generation/collection of data or subsequent addition of data to the same specification.

2.4 Dataset – A named collection of logically related features including processed data or information.

2.5 Geospatial Data – All data which is geographically referenced

2.6 Information – Processed data

2.7 Metadata – The information that describes the data source and the time, place, and conditions under which the data were created. Metadata informs the user of who, when, what, where, why, and

how data were generated. Metadata allows the data to be traced to a known origin and know quality.

2.8 Negative List – Non-sharable data as declared by the departments/ organizations

2.9 Restricted Data –Data which are accessible only through a prescribed process of registration and authorization by respective departments / organizations.

2.10 Sensitive Data– Sensitive data as defined in various Acts and rules of the Government of India.

2.11 Sharable Data – Those data not covered under the scope of negative list and non-sensitive in nature

2.12 Standards – Any application that embeds data handling functions (e.g., data collection, management, transfer, integration, publication, etc.) and operates on data in a manner that complies with data format and data syntax specifications produced and maintained by open, standards bodies.

Acts and rules of Government of India, thereby permitting a wider accessibility and use of public data and information.

17.3 Scope of this Policy

The National Data Sharing and Accessibility Policy will apply to all data and information created, generated, collected and archived using public funds provided by Government of India directly or through authorized agencies by various Ministries / Departments /Organizations / Agencies and Autonomous bodies.

17.4 Benefits of the data sharing policy

Maximizing Use – Ready access to government owned data will enable more extensive use of a valuable public resource for the benefit of the community.

Avoiding Duplication – By sharing data the need for separate bodies to collect the same data will be avoided resulting in significant cost savings in data collection.

Maximized integration – By adopting common standards for the collection and transfer of data, integration of individual datasets may be feasible.

Ownership information – The identification of owners for the principal datasets provide information to users to identify those responsible for implementation of prioritized data collection programs and development of data standards.

Better decision-making – Data and information facilitates making important decisions without incurring repetitive costs. Ready access to existing valuable data is essential for many decision making tasks such as protecting the environment, development planning, managing assets, improving living conditions, national security and controlling disasters.

Equity of access – Amore open data transfer policy ensures better access to all bonafide users.

17.5 Legal framework

Data will remain the property of the agency/department/ ministry/ entity which collected them and reside in their IT enabled facility for sharing and providing access. Access to data under this policy will not be in violation of any Acts and rules of the Government of India in force. Legal framework of this policy will be aligned with various Acts and rules covering the data.

17.6 Pricing

Pricing of data, if any, would be decided by the data owners and as per the government policies. All Ministries / Departments will upload the pricing policy of the data under registered and restricted access within three months of the notification of the policy. A broad set of parameters

would be standardized and provided as guidelines for the use of data owners.

17.7 Implementation

a) The Department of Science & Technology serving the nodal functions of coordination and monitoring of policy through close collaboration with all Central Ministries and the Department of Information Technology by creating https://data.gov.in through National Informatics Centre (NIC).

b) All sharable data will be made available on 'as-is where-is' basis.

c) Detailed implementation guidelines including the technology and standards for data and metadata would be brought out by Department of Information Technology, Government of India.

d) All the data users who are accessing/using the data shall acknowledge the ministry/department in all forms of publications.

e) All Ministries/Departments will upload at least 5 high value datasets on https://data.gov.in within three months of the notification of the policy.

f) Uploading of all remaining datasets should be completed within one year

g) Thereafter, all datasets are to be uploaded regularly every quarter.

h) https://data.gov.in will have the metadata and data itself and will be accessed from the portals of the departments/ministries.

i) The metadata in standardized formats is to be ported on https://data.gov.in which enables data discovery and access through departmental portals. All metadata will follow standards and will minimally contain adequate information on proper citation, access, contact information, and discovery. Complete information including methods, structure, semantics, and quality control/assurance is expected for most datasets.

j) Government will design and position a suitable budgetary incentive system for data owners for increasing open access to the sharable

data.

k) An oversight committee will be constituted for facilitating the implementation of the policy and its provisions thereof

l) Department of Information Technology will constitute a coordination committee for implementation.

17.8 Budget Provisions

The implementation of National Data Sharing and Access Policy is expected to entail expenditures for both data owners and data managers for analog to digital conversion, data refinement, data storage, quality up-gradation etc. Budgetary provisions and appropriate support for data management for each department / organization by Government of India would be necessary.

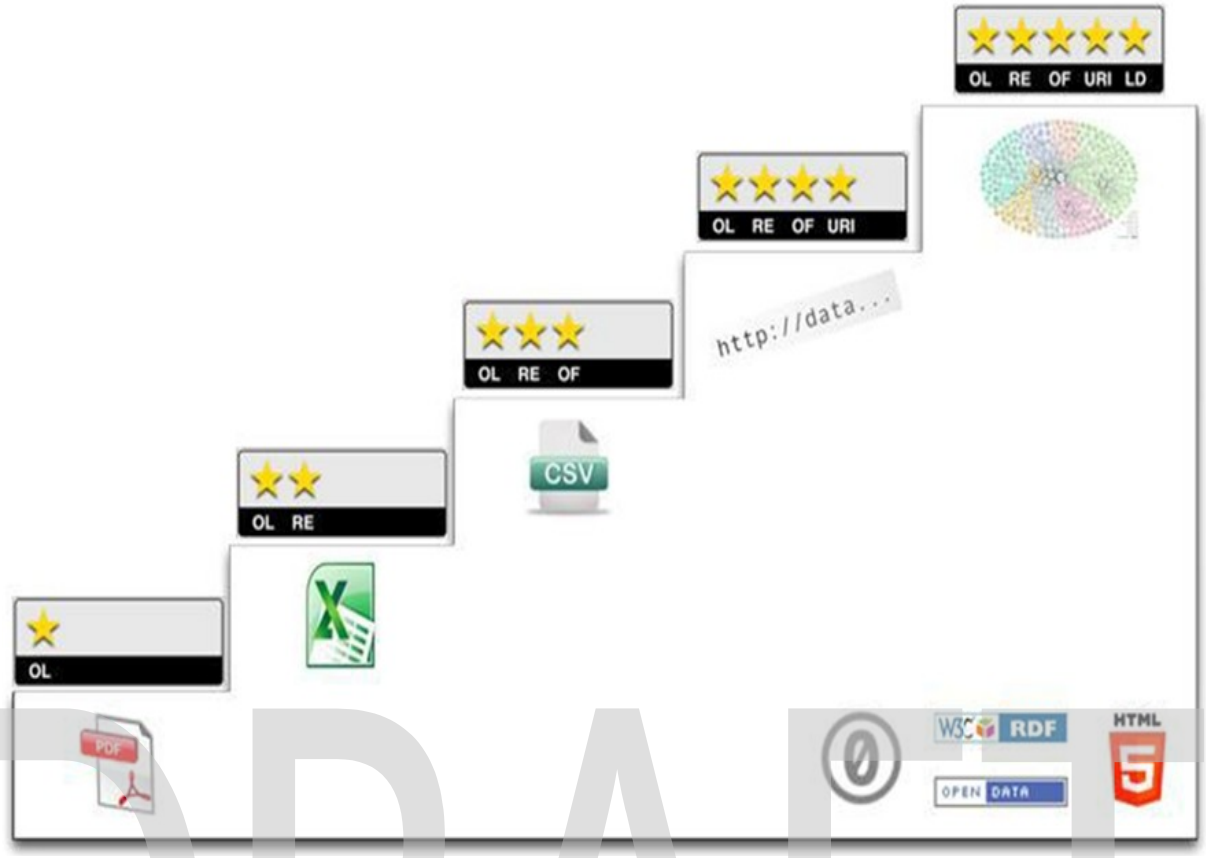
17.9 Conclusion

While policies provide official mandate, facilitation of optimum accessibility and usability of data by the implementers pre-suppose a trajectory of proper organization of data, with access services and analysis tools that provide the researchers and stakeholders with added value. For data to be reused, it needs to be adequately described and linked to services that disseminate the data to other researchers and stakeholders. The current methods of storing data are as diverse as the disciplines that generate it. It is necessary to develop institutional repositories, data centres on domain and national levels that all methods of storing and sharing have to exist within the specific infrastructure to enable all users to access and use it.

National Data Sharing and Access Policy aims at the promotion of a technology-based culture of data management as well as data sharing and access. It opens up, proactively, information on available data, which could be shared with civil society for developmental purposes, their price details if any, and methods for gaining access to registered and restricted use. The policy has limited its scope to data owned by the agencies, departments/ Ministries and entities under the Government of India and forms a statement of the Government of India of its commitment to transparency and efficiency in governance. Department of Science & Technology will continue the process of evolving the policy further, keeping in tune with technological advancements and the National requirements and enrolling the State Governments.

18. Five Star Open Data

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5 star deployment scheme for Open Data. Here, we give examples for each step of the stars and explain costs and benefits that come along with it.



Where the abbreviations stands for:

18.1 OL – Open License

18.2 RE – Readable (Human & Machine)

18.3 OF – Open Format

18.4 URI – Universal Resource Identifier

18.5 LD – Linked Data

The cost and benefit of these open data levels are detailed below:

★ **Available on the web (whatever format) but with an open licence, to be called as Open Data**

As a consumer....

- You can look at it.
- You can print it.
- You can store it locally (on your hard drive or on an USB stick).

- You can enter the data into any other system.
- You can change the data as you wish.
- You can share the data with anyone you like.

As a publisher...

- It's simple to publish.
- You do not have explained repeatedly to others that they can use your data.

Note: It's great to have the data accessible on the Web under an open license (such as PDDL, ODC-by or CC0), however, the data is locked-up in a document. Other than writing a custom scraper, it's hard to get the data out of the document

★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

As a consumer, you can do all what you can do with ★ Web data & additionally…..

- You can directly process it with proprietary software to aggregate it, perform calculations, visualise it, etc.
- You can export it into another (structured) format.

As a publisher...

- It's still simple to publish

Note: Splendid! The data is accessible on the Web in a structured way (that is, machine-readable), however, the data is still locked-up in a document. To get the data out of the document you depend on proprietary software.

★ ★★★ As (2) plus non-proprietary format (e.g. CSV instead of excel)

As a consumer, you can do all what you can do with ★★ Web data and additionally…..

- You can manipulate the data in any way you like, without being confined by the capabilities of any particular software.

As a publisher...

- You might need converters or plug-ins to export the data from the proprietary format.
- It's still rather simple to publish.

Note: Excellent! The data is not only available via the Web but now everyone can use the data easily. On the other hand, it's still data on the Web and not data in the Web.

★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff using an URI

As a consumer, you can do all what you can do with ★★★ Web data and

Additionally...

- You can link to it from any other place (on the Web or locally).
- You can bookmark it.
- You can reuse parts of the data.
- You may be able to reuse existing tools and libraries, even if they only understand parts of the pattern the publisher used.
- Understanding the structure of an RDF "Graph" of data can be more effort than tabular (Excel/CSV) or tree (XML/JSON) data.
- You can combine the data safely with other data. URIs are a global scheme so if two things have the same URI then it's intentional, and if so that's well on it's way to being 5 star data!

As a publisher...

- You have fine-granular control over the data items and can optimise their access (load balancing, caching, etc.)
- Other data publishers can now link into your data, promoting it to 5 star!
- You typically invest some time slicing and dicing your data.
- You'll need to assign URIs to data items and think about how to represent the data.
- You need to either find existing patterns to reuse or create your own.

Note: Wonderful! Now it's data in the Web. The (most important) data items have a URI and can be shared on the Web. A native way to represent the data is

using RDF. However, other formats such as Atom can be converted/mapped, if required.

★ **All the above, plus: Link your data to other people's data to provide context**

As a consumer, you can do all what you can do with ★★★★ Web data and additionally…..

- You can discover more (related) data while consuming the data.
- You can directly learn about the data schema.
- You now have to deal with broken data links, just like 404 errors in web pages.
- Presenting data from an arbitrary link as fact is as risky as letting people include content from any website in your pages. Caution, trust and common sense are all still necessary.

As a publisher…

- You make your data discoverable.
- You increase the value of your data.
- You own organisation will gain the same benefits from the links as the consumers.
- You'll need to invest resources to link your data to other data on the Web.
- You may need to repair broken or incorrect links.

Note: Brilliant! Now it's data, in the Web linked to other data. Both the consumer and the publisher benefit from the network effect

19. Ten Principles for Opening Up Government Information

In October 2007, 30 open government advocates met in Sebastopol, California to discuss how government could open up electronically stored government data for public use. Up until that point, the federal and state governments had made some data available to the public, usually inconsistently and incompletely, which had whetted the advocates'

appetites for more and better data. The conference, led by Carl Malamud and Tim O'Reilly and funded by a grant from the Sunlight Foundation, resulted in eight principles that, if implemented, would empower the public's use of government held data.

Sunlight Foundation has updated and expanded upon the Sebastopol list and identified **ten principles** that provide a lens to evaluate the extent to which government data is open and accessible to the public. The list is not exhaustive, and each principle exists along a continuum of openness. The principles are completeness, primacy, timeliness, ease of physical and electronic access, machine readability, non-discrimination, use of commonly owned standards, licensing, permanence and usage costs.

19.1 Completeness

Datasets released by the government should be as complete as possible, reflecting the entirety of what is recorded about a particular subject. All raw information from a dataset should be released to the public, except to the extent necessary to comply with federal law regarding the release of personally identifiable information. Metadata that defines and explains the raw data should be included as well, along with formulas and explanations for how derived data was calculated. Doing so will permit users to understand the scope of information available and examine each data item at the greatest possible level of detail.

19.2 Primacy

Datasets released by the government should be primary source data. This includes the original information collected by the government, details on how the data was collected and the original source documents recording the collection of the data. Public dissemination will allow users to verify that information was collected properly and recorded accurately.

19.3 Timeliness

Datasets released by the government should be available to the public in a timely fashion. Whenever feasible, information collected by the government should be released as quickly as it is gathered and collected. Priority should be given to data whose utility is time sensitive. Real time information updates would maximize the utility the public can obtain from this information.

19.4 Ease of Physical and Electronic Access

Datasets released by the government should be as accessible as possible, with accessibility defined as the ease with which information can be obtained, whether through physical or electronic means .Barriers to physical access include requirements to visit a particular office in person or requirements to comply with particular procedures (such as completing forms or submitting FOIA requests). Barriers to automated electronic access include making data accessible only via submitted forms or systems that require browser oriented technologies (e.g., Flash, JavaScript, cookies or Java applets). By contrast, providing an interface for users to download all of the information stored in a database at once (known as "bulk" access) and the means to make specific calls for data through an Application Programming Interface (API) make data much more readily accessible. (An aspect of this is "find ability," which is the ability to easily locate and download content.)

19.5 Machine readability

Machines can handle certain kinds of inputs much better than others. For example, hand written notes on paper are very difficult for machines to process. Scanning text via Optical Character Recognition (OCR) results in many matching and formatting errors. Information shared in the widely used PDF format, for example, is very difficult for machines to parse. Thus, information should be stored in widely used file formats that easily

lend themselves to machine processing. (When other factors necessitate the use of difficult to parse formats, data should also be available in machine friendly formats.) These files should be accompanied by documentation related to the format and how to use it in relation to the data.

19.6 Non-discrimination

"Non-discrimination" refers to who can access data and how they must do so. Barriers to use of data can include registration or membership requirements. Another barrier is the use of "walled garden," which is when only some applications are allowed access to data. At its broadest, non-discriminatory access to data means that any person can access the data at any time without having to identify him/herself or provide any justification for doing so.

19.7 Use of Commonly Owned Standards

Commonly owned (or "open") standards refer to who owns the format in which data is stored. For example, if only one company manufactures the program that can read a file where data is stored, access to that information is dependent upon use of the company's processing program. Sometimes that program is unavailable to the public at any cost, or is available, but for a fee. For example, Microsoft Excel is a fairly commonly used spread sheet program which costs money to use. Freely available alternative formats often exist by which stored data can be accessed without the need for a software license. Removing this cost makes the data available to a wider pool of potential users.

19.8 Licensing

The imposition of "Terms of Service," attribution requirements, restrictions on dissemination and so on acts as barriers to public use of

data. Maximal openness includes clearly labelling public information as a work of the government and available without restrictions on use as part of the public domain.

19.9 Permanence

The capability of finding information over time is referred to as permanence. Information released by the government online should be sticky: It should be available online in archives in perpetuity. Often times, information is updated, changed or removed without any indication that an alteration has been made. Or, it is made available as a stream of data, but not archived anywhere. For best use by the public, information made available online should remain online, with appropriate version tracking and archiving over time.

19.10 Usage Costs

One of the greatest barriers to access to ostensibly publicly available information is the cost imposed on the public for access—even when the cost is de minimus. Governments use a number of bases for charging the public for access to their own documents: the costs of creating the information; a cost recovery basis (cost to produce the information divided by the expected number of purchasers) the cost to retrieve information, a per-page or per-inquiry cost, processing cost, the cost of duplication etc.

Most government information is collected for governmental purposes, and the existence of user fees has little to no effect on whether the government gathers the data in the first place. Imposing fees for access skews the pool of who is willing (or able) to access information. It also may preclude transformative uses of the data that in turn generates business growth and tax revenues.

20. Implementing Dharamshala Data Smart City

Data Smart Cities will be implemented in consonance with NDSAP Guidelines. Following institutional structure will support its implementation.

Mission Director, SCM, will nominate a Mission Data Officer (MDO) for SCM. The MDO will be an officer not below the rank of a Director in MoHUA. The MDO will be the officer responsible for implementation of the strategy at the national level. The MDO will engage with all the 100 smart cities, NDSAP Nodal Ministry (MEITY), NIC and other relevant agencies and organisations to achieve the outcomes outlined under the strategy. The MDO will continuously identify key data sets and high value data feeds to be published on the OGD Portal. The MDO will prepare and continuously review a negative list in line with NDSAP and other government policies in consultation with SCDOs.

1. A Mission Data Hub (MDH) will be constituted within the Smart Cities Mission office. The MDH cell will function as Project Management Unit (PMU) to support the MDO in implementing Data Smart Cities strategy. The cell would be constituted of relevant experts in legal frameworks, data science, data analytics, communications and other relevant fields. The cell would act as key support structure to the MDO to evolve the strategy over time, create capacity building within the smart cities ecosystem, coordinate with different stakeholders, advise on legal frameworks, and create data analytics capabilities within the mission. The MDH Cell will provide coordination, implementation, monitoring and hand holding support to Smart Cities. The MDH cell will review progress with smart cities every month based on defined KPIs, chalk out plans, share ideas, brainstorm new use cases, enable peer-to-peer learning and build and share solutions around data sharing, privacy, exchange, data driven governance amongst various stakeholders of the smart cities and release a status report on the implementation of the strategy every quarter. In short, the MDH cell

would be the backbone for implementation of strategy at the national level.

2. Municipal Commissioners of Smart cities will designate a senior official of the rank of a Chief Technology Officer/Chief Information Officer/Department Head IT or above as the City Data officer (CDO). All smart cities will designate their CDOs within 15 days from the announcement of this strategy. The CDO will be the officer responsible for implementation of this data strategy at the city level. The MDO at the national level, and the CDOs in the cities will continuously identify key data sets and high value data feeds to be published on the OGD Portal. High value data sets are those data sets, which create large scale empowerment of communities and enable transformative data driven governance and co-creation possible in areas which are most critical to the city's needs.
3. Smart Cities Data Network (SCDN) will be a network of selected CDOs from 100 smart cities, along with representatives from other Ministries of Government (both State and Central), industry associations, research organizations, academic institutions and legal firms amongst others. The network will be termed as. MDO will be the convener of the SCDN. The SCDN will meet at least once every quarter, physically or virtually. SCDN will act as an advisory body for Data Smart Cities strategy of SCM. SCDN will help define policy contours, bring in expertise in understanding the data landscape in other parts of the world, identify best practices in Indian cities, help in interpreting the legal framework around data. SCDN, thus will help the Mission in its effort to use data as a tool for empowerment of the society.
4. City Data Alliance (CDA) comprising of various key stakeholders including city government, other key actors in policy making, various government department and agencies, representatives of leading academic and research institutions in the city, community organizations, entrepreneurs and advocacy groups should be created in the cities. The Municipal Commissioners of smart cities will facilitate

the creation of active CDAs in their cities. The CDO will be the convener of the CDA.

5. Data-Smart Cities Strategy is applicable to all 100 Smart Cities under Smart Cities Mission of Ministry of Housing and Urban Affairs.

21. Data Governance

21.1 Mission Data Officer (MDO)

The Key responsibilities of MDO are as follows:

1) MDO will manage MDH cell at SCM Office under MoHUA.
2) MDO will work with Smart City Data Officers (SCDOs) of 100 Smart Cities.
3) MDO will provide coordination, implementation, monitoring and hand holding support to all SCDOs and will work as the convener for the Smart Cities Data Network (SCDN). Refer section 2.1 for more details on SCDN.
4) MDO will prepare the Mission Data Strategy and Plan (MDSP).
5) MDO will, along with the SCDOs, identify and suggest the list of data sets/feeds to be published by Smart Cities on Data Platform. The list of data sets/ feeds so prepared should be reviewed every month by the MDO in consultation with CDOs.
6) MDO will engage with CDOs, NDSAP Nodal agency (MEITY), NIC and other relevant agencies, to achieve the Mission goals outlined under Mission Data Strategy.
7) MDO will monitor the Key Performance Indicators as set by MDH cell and will prepare regular dashboard and report to assess progress as per Mission Data Strategy and Plan.
8) MDO will engage with the Smart Cities Stakeholders to identify the data needs and demand on regular basis.
9) MDO will guide SCDOs in ensuring security and privacy aspect as per NDSAP and other prevailing laws.

10) MDO will review the data quality of data available under SCM on Data portal.
11) MDO will publish white paper every quarter to bring forth the overall progress in implementation of the strategy.
12) MDO will formulate data taxonomy for data sets/feeds. The SCDOs will refer the data taxonomy to publish data sets/feeds.
13) MDO will conduct data challenges, maturity benchmarking of cities on continuous basis with the help of MDH.

21.2 City Data Officer (CDO)

The Municipal Commissioners of Smart cities will designate a senior official of the rank of a Chief Technology Officer/Chief Information Officer/ Department Head IT or above as the City Data officer (CDO). The CDO will act as custodian and driver of City Data plan (CDP) and a flag bearer of open government initiative in respective city. CDO's major responsibility is to put data to its right use i.e. for generating insights, using data for effective service delivery or infrastructure delivery, improving civic operations by making real time decision making etc. City data officer will work with city leadership to assess and tap the potential of data and set up data culture across the organization and outside the organization.

The key responsibilities of CDO are as follows:

1) The CDOs will create a City Data Policy (CDP) for their respective smart cities which will be reviewed every month to keep it contextual to the need of the times. The first City Data Policy shall be prepared by the Smart Cities within two months from announcement of this strategy. The policy should be created post engagement with relevant stakeholders. The CDA would act as advisory body for the review of CDP from time to time. It will be responsibility of Municipal Commissioner to ensure that the policy

evolves as per the needs of various stakeholders of the city and relevant upgrades to policy are carried out time to time accordingly.

2) Coordinate with MDO to align with mission data strategy and priorities with respect to Open government initiatives and policies.
3) Organise regular meetings of CDA.
4) Coordinate with officers of various other government departments/agencies within the city for the effective implementation of City Data Policy.
5) Publish Data Catalogues and Data Sets/Feeds on OGD portal: CDOs will publish data Catalogues and Data Sets/Feeds on OGD Portal and will ensure that such data sets are updated at regular time intervals as needed and create mechanisms for continuous feedback from citizens and stakeholders on type of data sets to be published.
6) The CDO will be responsible for publishing of such data sets/feeds as mandated as part of Mission Data Strategy.

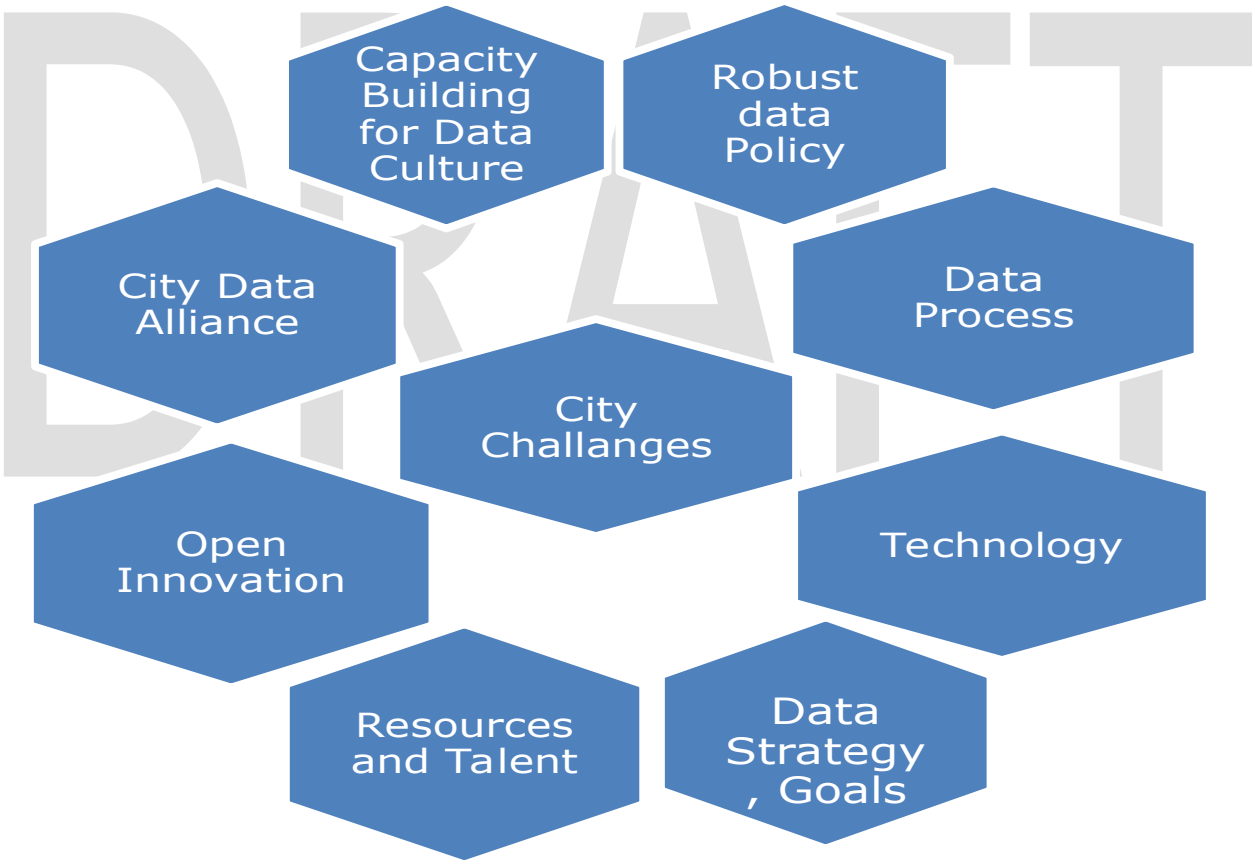
Tips: City Leadership needs to provide full administrative and logistic support to CDO. This only can make possible a city wide change in data culture. Leadership commitment and communication will be a crucial aspect in achieving goals highlighted under the strategy Thus, it is advised that City leadership active interest in appointing suitable CDO and providing full support to him/her to successfully implement the City Data Strategy. City Data officer will report directly to City Leadership and act as single point of contact to all internal and external stakeholders in the city. Leadership need to also deploy dedicated skilled resources to drive the data initiative through City Data Officer. Integrating right talent with required business knowledge and technical skill set is key requirement for driving the city data office. City Leadership need to assign dedicated budget to on-board resources for City Data Officer. Apart from hiring resources, dedicated funding should be committed for capacity building, change management, setting up open data platform and knowledge sharing.

Core objective of setting up City Data Office is to focus on setting up data driven governance culture across organization. It is implied that city leaders will be providing the required leadership support to drive the data driven decision making through seamless data collection, processing and analysis across all departments/government agencies.

22. Key Building Blocks of Data Culture

Driving the Data Culture Transformation

In order to create a self-sustaining, usable data management ecosystem various elements will be required to come together so true value may be derived from the data being captured



Strategy Formulation and Mind-set Transformation – The first step towards building a sustainable data culture is to identify current data goals, to strategize/pre-empt any upcoming data priorities. What are the

final outcomes that are required from the data? What sort of value needs to be added from the analysis of said data? Who all will be using the data? How will the data be accessed?

The only way a sustainable 'Data-Culture' can be created if the mind-set of users from officers to last mile data collectors is in sync with the importance of collecting the data and the value that it adds by providing varied actionable insights.

Although the chosen technological platforms may be conducive to allow easy collection and management of data, re-iterating the importance of messages would go a long way in creating a data driven culture.

1. **Robust Data Policy –** The aim should be to build a Safe, Secure, Easily Accessible environment for data collection, collation and analysis. The data governance policy should serve as clear documented set of guidelines for ensuring the proper management of managing digital information across all stages of the data management lifecycle.
2. **Processes –** Processes need be formulated in a way so as to collect data in an efficient manner. Data collected should be accurate, clean and be supported by layered reviewing capability basis importance of data and available resources in the organisational structure chain.
3. **Technology -** The technology should be future ready to manage the pre-empted/strategized data needs.

   Open Source and Proprietary technologies that provide tremendous data capture and analysis capabilities are readily available. These should be chosen bearing in mind factors like data usage goals (data indexing, depth of analytics required), storage and indexing needs, technical expertise available and financial appetite.

   In most cases a lot of data is already available with various government bodies that could of use to the Smart City. The

chosen tools and corresponding processes should also be able to cater to the data demands from various entities.

It should be noted that ease of data collection should be the top priority at data collection points since the users collecting the data may not be technological sound in some cases. Data should be secure and accessible according to the user rights both in raw or analysed form. The analysed data should be accessible on multiple devices and other desired channels.

4. **Capacity Building –** Post identification of tools and formulation of processes, users at all stages will require training on how to collect data and using the analytics tools.

   After the technology intervention and processes have been identified, focus on capacity building, technical training will happen.

   At times the resources deployed at the last mile may not understand the value of meticulous and focussed data collection and may need to be trained not only on the tools but also on the importance of collecting data properly.

5. **Data Alliances –** While sustainable data influx and can be ensured using above building blocks, it is importance not to ignore already present data at various other government and private sources. Such data can help serve as an important supplement to the existing data sources and on channels as desired.

   This data might not be captured in the same format and will need cleaning and normalization before analysis can be done.

6. **Resource Building –** In order to handle the complexities of the data, it is important that the right resources be deployed at critical points in the data life cycle. These resources can serve to guide the teams at these critical points in the data life cycle so as to avoid spill overs, incorrect data entry, cleansing techniques etc. These data scientists/data champions will be instrumental in

generating useful insights and deliver value to the senior stakeholders.

7. **Open Innovation –** Innovation is an important element in the success of any technological initiative. In order to drive innovation Hackathons/Innovation competitions could be held. Better techniques tailored to the Smart City needs could be used or useful data could be shared with participants to drive new insights that may have been previously missed.

23. Data Champions and Coordinators

The CDO cannot work in isolation. Active participation from data agencies will be key to successful data collaboration within the city. Data champions will be senior functionaries, not below the rank of a Head of Department or equivalent, who would champion the implementation of the CDP in their respective departments/ organizations. They would be the flag bearers of the policy in their departments/ organizations and would work to align their teams to imbibe the principles of data driven decision making in their day to day functioning. They would also make their respective teams aligned to the value of collaborative work on data, as siloed approach to data ownership and use will not foster the development of integrated approaches to resolution of potential use cases within the context of each city.

24. Data Champions (DC):

Data Champions (DCs) in respective departments/government agencies shall identify the data sets/feeds, derived information, intelligence or data challenge with respect to day to day operations of the department.

1) DCs will actively publish/ enable to publish data sets/feeds identified as relevant to the resolution of critical use cases for the city. They will work closely with the CDO for active implementation of the City Data Policy.

2) DCs will be assisted by the Data Coordinators (DCs) within the department to streamline processes of data reporting, collection and analysis etc. Data Champions will be responsible for data quality.
3) DCs will undertake activities to engage with their stakeholders and evolve their department's strategy on data in line with the deliberations.

Tips: DCs needs to act as trainers and lead the team of data coordinators at the department level. It is critical for appropriate senior functionaries to be designated as DCs in each department by the respective city heads. The DC will the nodal point for implementation of the CDP within the department and will function to supervise the team of data coordinators on day to day basis. DC will be first touch point of CDO in different city organizations and must undertake continuous capacity building programs for their DCOs and other staff.

25. Data Coordinators:

1) Data Coordinators will assist DCs at the department/government agency level as reporting staff.
2) Data Coordinators will also aggregate the data demand from various channels.
3) Data Coordinators will also be responsible for sensitizing the department employees over the importance of data quality etc.

26. City Data Alliance (CDA)

A CDA for a city is envisaged to be network of government departments, agencies, private sector companies, community organizations, city policy makers, domain & legal experts, research, academic institutions, incubators, entrepreneurs, etc, within the city who come together voluntarily as a collective to diagnose city problems which need resolution, act as an advocacy group for the formulation of the city data policy (CDP) which defines the collective approach of the city on issues related to data. The CDA will provide a collaborative framework to create

and define use cases to solve critical city problems through the use of data, catalyse the right set of collaborations and networks to make available such data and undertake continuous dialogue between various stakeholders in the city around the City Data Policy so as to inform and evolve the CDP effectively. Thus, a CDA is not merely an advocacy organization or a think tank, but it is a network of organizations who not only understand the value of data towards improving quality of life, employability and sustainability in the city, but also be the chief driver of adoption of data driven governance and co-creation within the community. The alliance will undertake education and awareness about data in the community, understand and address concerns on data privacy and security, build use cases for city problems, create data collaborations between various government and private agencies for solving relevant use cases and continuously evolve the culture of data in the city's context.

26.1 Roles of City Data Alliance (CDA):

The key objectives for setting up City Data Alliance (SCDA) are as follows:

a) To act as an advisory group to the city leadership on the City Data Policy.
b) To assess the data needs of various Smart City stakeholders.
c) To promote data driven governance and policy formulation.
d) To design and implement solutions and analysis using city data.
e) To support industry to design solutions using emerging technologies like AI, ML and Block chain.
f) To assess and design use cases critical to the citizens of the respective cities.
g) To generate awareness in various stakeholders towards open government initiatives.
h) To bring Smart Cities stakeholders on common platform to influence the city data priorities.
i) To facilitate data for co-creation and collaboration over civic issues

j) To provide critical feedback to the city over the quality and relevance of data provided by Smart City.

k) To deliver 4 Research paper annually using City Data on Civic Problems in Smart City.

l) To design and develop two prototype/ solutions annually on Civic Problems in Smart City.

m) To organize a data challenge every half yearly on complex civic problems.

n) To organize a Hackathon annually and support shortlisted solutions at city level.

o) To set up scholarship for postgraduate and graduate interns to work with Office of CDO.

p) To publish the progress report every month.

q) Prioritize the Data Sets/Feeds for publishing on Data Platform.

r) To sensitize ecosystem partners to share the data for leveraging data for solving civic challenges

s) To support, engage and encourage network/groups/members of data enthusiasts in Smart City

t) To improve city capacity over data driven governance and policy formulation

u) To support SCDO by extending resources (like interns, researchers, technology experts), funds (program sponsorship etc.) and technology (solutions etc.)

v) To share data available with partners on Data Platform to promote City Data.

26.2 Need for city Stakeholders to come together in CDA:

Following stakeholders should come together to set up City Data Alliance to assess, strategize, plan , implement and review the City Data Policy :

a) **Government Agencies:** Government Agencies operating with dedicated administrative structure (apart from City Administration) in Smart City namely Traffic Police, City Police, Central/State Government Departments, Government Autonomous Bodies etc.

b) **Funding Agencies:** Funding Agencies which regularly works with city administration in different domain for e.g. World Bank, ADB, DFID, AFB etc.

c) **Industry:** Key flagship manufacturing/service Industry promoters/ players in the Smart city/state.

d) **Academia:** Representatives from leading Universities/Colleges /Schools in the Smart city.

e) **Policy Advocacy Groups and NGOs:** Policy Advocacy groups and NGOs working in different domains/areas like Slums, Health, Education, Environment, Participatory Governance, Mobility etc.

f) **Start-ups and Incubators:** Representatives from start-ups and incubators in the Smart City/State.

g) **City Businesses:** Representatives from local Small and medium business communities

h) **Citizens and Communities:** Representatives from Communities and citizen interest groups to further the interest of citizens/communities towards data driven policy governance and service delivery.

i) **Local Elected Representatives:** Representatives from local elected representatives to further the interest of citizens/communities towards data driven policy governance and policy formulation.

j) **Professional Representatives:** Representatives from various professional bodies like Doctors, CA, and Engineers etc.

27. Smart Cities Data Network (SCDN)

The Smart Cities Data Network (SCDN) is conceived as a national forum that will provide guidance related to the implementation of this strategy to the office of the MDO and CDOs of all smart cities. The SCDN will be a national forum which will have as its members representatives of CDAs from few Smart Cities chosen on rotation basis from year to year, some CDOs from the cities, Municipal Commissioners from a few cities, representatives from various Government departments/ agencies from both the Central/ State levels, academia and research organizations, industry associations, experts and other key members from the Government/ private sector actively working in fields related to data governance/privacy/management etc.

SCDN will act as an advisory group to the office of the MDO and will contribute actively to evolve the Mission Data Strategy. The recommendations of SCDN would become inputs for policy formulation around data at the National level. The CDA in each city will the alliance at the city level translating national policies and guidelines to a locally meaningful manner, and providing city level feedback and requirements to the SCDN through CDO.

27.1 SCDN Activities and Responsibilities

a) Advise the MDO on Data Policy guidelines, and recommendations in accordance with NDSAP and other relevant laws in this regard.

b) Advice on capacity building around data in Smart Cities.

c) Extend advisory support in evaluating the functioning of City Data Alliances (CDAs) and recommend suitable course of action as per observations, so as to strengthen the working of CDAs in each of the Smart Cities.

d) Provide advisory support to the MDO over various relevant policies/ laws/ rules and regulations about data sharing/ privacy/ management etc.

e) Provide advice on organization of key events like hackathons, conferences, etc.

f) Advice on global best practices on data for cities.

Tips: Set up desk or online mechanism to understand the data needs of various Smart Cities Stakeholders. Engage the stakeholders and address the concern related to data accessibility, data availability, data quality etc. Also set up discussion forums etc. to gauge the community sentiment or feedback on data initiatives.

28. Dharamshala Smart City Data Policy

- Data Classification – YES
- Data Categorization – YES
- Data Flow & Approval Framework – DFD and approval hierarchy pending
- Data Archieval and Retention – Sizing pending
- Data security – pending
- SoP for data collection
- SoP for electronic data collection
- SoP for data processing and cleaning
- SoP for data for quality assessment of Data sets
- SoP for data publishing as per Open Data Norms
- SoP for engaging stakeholders to assess the data needs – Pending for PIU mtg
- SoP for data collection, processing and analysis for on field Survey – Make DFD.
- Provisions for Data Analysis

29. References

1. Source: From Wikipedia
2. Source from: Main article: data.gov.in

3. Source: http://sunlightfoundation.com/policy/documents/ten-open-data-principles/
4. Sources and References: Gazette Notification for NDSAP can be accessed at: https://data.gov.in/sites/default/files/NDSAP.pdf
5. (Source: http://5stardata.info/)