IMPLEMENTATION GUIDELINES OF CITY DATA POLICY

FOR AGARTALA CITY

Scanned by CamScanner

Overview

A Smart City is a city where "people and businesses are empowered through increased access to data, more participatory through the contribution of innovative ideas and solutions, and a more anticipatory government that utilises technology to better serve citizens. This City Data Policy enables greater pervasive connectivity, better situational awareness through data collection, and efficient sharing and access to collected sensor data, allowing public bodies to use such data to develop policy and practical interventions. Such access would allow for anticipatory governance - a goal of the City Data Policy stating "Insights gained from this data would enable us to better anticipate citizens' needs and help in better delivery of services".

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 City Data Policy 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.

Background

Cities around the world are confronted with a range of challenges. These include demographic change and the rising cost of healthcare, traffic congestion and high levels of emissions, the need to improve employment prospects and the quality of work in the face of technical change and, following the financial crisis, to deliver services with significantly reduced budgets and higher citizen expectations. An increasing number of cities are turning to "smarter" approaches in planning their future, in creating a modern infrastructure and in delivering services. Government of India has played an important role in shaping thinking in how innovation can be harnessed to improve cities, particularly in the development of national and international standards.

The Objectives of the Data Smart Cities

- **Institutionalize a "Culture of data":** The trend to draw insights and create actionable intelligence for city governance is already on the rise. However, formal mechanisms for data collection, management and use needs to be put in place. Also, there is a need to create awareness, dialogue and collaboration among different stakeholders to harness the power of data as a potential economic resource.

- **Drive Data Governance:** To propose a data governance framework that facilitates the implementation of key processes within the data life cycle and builds capacity in

Scanned by CamScanner

all stakeholders on data- informed decision-making. The data strategy should foster public accountability and transparency.

• **Enable the framing of a City Data Policy:** To unlock the power of data in the context of privacy, security and ownership in the context of the city, it is critical that cities create data policies that balance privacy, legal and public benefit considerations. At the same time it must define the contours of collaboration between various Governmental/non-Governmental entities on data sharing and access.

• **Facilitate City Data Alliance:** It is important to assess the data available in all Government and non- Government entities that generate and store data crucial to better planning and functioning of the city and to engage them in the understanding, creation, and promotion of data-driven solutions for the city. The 'quadruple helix' comprising of communities, industry, academia and the Government will be constituents of the City Data Alliance.

• **Adopt appropriate Data platforms:** With a clear strategy, cities can adopt and deploy robust, secure and intuitive data exchange platforms, which will lead to the effective sharing and management of city data. Such platforms allow common programming interfaces, data representation formats and data models that are interoperable.

Benefits of the Data Smart Cities Strategy

The implementation of this initiative in Smart Cities will lead to the following benefits:

• Empowerment of citizens: When cities are open about how they function, connect with their through various platforms during the development lifecycle of myriad projects, put out information about their tax collections, their financial and environmental health and are open to informed debates, they become true proponents of 'Open Government'. Such cities constantly try to build trust with their citizens and engender a collective conscience amongst their communities around important issues facing their present and future. Citizens can collaborate with government easily and with increased frequency; both within their community and beyond, forming stronger groups and exchanging ideas and building new collaborations.

• Data-driven governance and policy formulation: Data will help City administrators in making better policies and decisions for the city. Data empowers city officials, citizens, and communities and helps promote evidence-based decision making. This will lead to greater efficiency in service delivery and resource allocation.

• Promotion of Data Sharing and Exchange: Efficient governance requires ready availability of relevant data. Unfortunately, data is locked up in various systems with different data owners. Open Data initiatives, data sharing, and exchange platforms will

Scanned by CamScanner

assist in facilitating G2G, G2C and G2B data sharing and exchange of data for effective decision making in real time.

• Promotion of Multi-disciplinary research on Civic Issues: Local data could unlock research on civic issues like transport, traffic and solid waste. Multidisciplinary researchers may provide different perspectives or solutions on civic issues to the city administration.

• Co-Creation and Open Innovation and Civic Engagement: City Governments will be able to work with entrepreneurs, industry, and academia to promote participation in governance, co-creation and open innovation. This will enable greater civic engagement through directed partnerships and

Policy Implementation and Monitoring

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

a. Nominate Chief Data Officer

b. Chief Data Officer in turn Nominate Data Contributors

c. Co-ordinate with inter departmental/governmental offices

d. Identify Datasets

e. Publish Catalogs and Resources (Datasets/Apps) on OGD Platform India.

Smart City Data Alliance

The SCDA 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. The alliance will undertake education and 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.

Scanned by CamScanner

Responsibilities of Smart City Data Alliance (SCDA)

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 Blockchain.

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 CDOs by extending resources (like interns, researchers, technology experts), funds (program sponsorship etc.) and technology (solutions etc.)

Scanned by CamScanner

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

Responsibilities of Chief Data Officer (CDO)

The CDO will act as custodian and driver of City Data Policy (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.

City Data officer will report directly to City Leadership.

a. 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 policy should be created post engagement with relevant stakeholders. The SCDA (Smart city Data Alliance) would act as advisory body for the review of CDP (City Data Policy) 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.

b. Coordinate with MDO (Mission Data Officer) to align with mission data strategy and priorities with respect to Open government initiatives and policies.

c. Organize regular meetings of Smart City Data Alliance (SCDA).

d. Coordinate with officers of various other government departments/agencies within the city for the effective implementation of City Data Policy.

e. 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.

The CDO will be responsible for publishing of such data sets/ feeds as mandated as part of Mission Data Strategy. 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 CDO. 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.

Scanned by CamScanner

Role and Responsibilities of Data Contributor (DC)

In order to cater to the contribution of the datasets from offices/organization under the Ministries/Departments, the Data Controller 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. The contributed datasets would be approved by the Chief Data Officer as the case may be.

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:

a. Responsible for ensuring quality and correctness datasets of his/her unit/division

b. Preparing and contributing the catalogs and resources along with the metadata on the OGD Platform.

c. Aggregate the data demand from various channels.

d. Sensitizing the department employees over the importance of data quality etc

Data Classification

Data will be classified into the following categories:

Level 1 Public Data available for public consumption and use.

Level 2 Internal Use Information which can only be disclosed to municipal corporation employees for managing operations or delivery of public services on a day to day basis.

Level 3 Sensitive Data regulated by any city/ State/Central law or regulation like privacy law etc.

Level 4 Protected Data which needs to be protected e.g. Identity of citizens and disclosure /notification needs to be issued by the municipal corporation in case of any breach or loss of data.

Level 5 Restricted Data which could lead to a threat to life or loss of public assets or critical infrastructure and are accessible only through a prescribed process of registration and authorization by respective departments.

Scanned by CamScanner

Data Categorisation

Data will be categorized into two broad categories:

Personal Data: Personal data means data consisting of information which is related to an individual who can be identified from that information (or from that and other information in the possession of the data users), including any expression of opinion about the individual but not any indication of the intention of the data user in respect to that individual. 'Data' is defined as information recorded in a form in which it can be processed by equipment operating economically in response to instructions given for that purposes.

Note: Personal Identifiable Information cannot be published by the City on Data Platforms under any data sets. Datasets must be anonymized before publishing.

Non-Personal Data: Non-personal data also refers to anonymous information/data, namely information which does not relate to an identified or identifiable natural person, or personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. Anonymization means excluding any personal identifiers from data sets.

Data Security and Privacy

Managing security and privacy of data is crucial to building and maintaining trust between ecosystem participants and thus will be a critical element of the city data policy. Data collection, sharing and analysis must be ring-fenced by a privacy first approach to guarantee protections for residents and users. Smart Cities should develop ethical frameworks for data ownership and privacy which overcome any gaps in current legislation.

The usage rules for data elements must specify for what purposes the data can or cannot be used. For example, the patient's name in a hospital record may be fine to use by a doctor for a treatment plan, but not for any analysis or marketing without the consent of the data owner (the patient).

For management of Privacy and Security of Data, it is recommended that all data access must be through Application Programming Interface (API) calls to ensure appropriate security controls. Cities should establish or comply with existing standards and certifications for data privacy and security. Except for open data, it is recommended that direct access to data be prohibited and use of APIs mandated. Data dissemination should be only to authenticated and authorized stakeholders (both internal and external) through data fiduciaries.

Scanned by CamScanner

Data Flow and Approval Framework

Data governance isn't something to set and forget; it's a strategic approach that needs to evolve over time in response to new opportunities and challenges. Therefore, a successful data governance framework has to become part of the organization's culture but such a shift requires listening – and remembering that it's about people, empowerment and accountability.

In most cases, a new data governance framework requires people – those in IT and across the business, including risk management and information security – to change how they work. Any concerns they raise or recommendations they make should be considered. You can encourage feedback through surveys, workshops and open dialog.

Once input has been discussed and plan agreed upon, it is critical to update roles and responsibilities, provide training and ensure ongoing communication. Many organizations now have internal certifications for different data governance roles who wear these badges with pride.

A top-down management approach will get a data governance initiative off the ground, but only bottom-up cultural adoption will carry it out.

SOP for Data Collection

There are two broad types of data sources. One is personal data and other is non-personal data (**described in Data classification and Categorisation**). Data will be collected from different Government Departments, Public Agencies, Private Academia by the Data Contributors in collaboration with Head of the Departments, of the various Government Offices, viz, Sub- divisional Offices, Office of the Executive Engineers, Urban Local Bodies and its subsidiaries (ward offices), e-governance cell etc. Data, then will be sent over to City data Officer for further actions.

The municipal commissioner has to issue memorandum/ order to the various departments (Government/non-Government) to coordinate with the data contributors for collection of data as per given format or whichever is applicable.

SOP for Data Processing and Cleansing

1.. Remove formula from the dataset.

2. Unmerge the merged cells and repeat the cell value in row and column headers (if required) not in data matrix.

3. Keep heading/variable name in the first row.

Scanned by CamScanner

4. Replace blank cells with NA.

5. Remove special characters from Data points of data matrix to make it machine readable and need to write a note (combination of row and column variable name) for removed special character in metadata note.

6. For each special character removed from any cell in data matrix, give proper referencing to that cell by mentioning the row and column (e.g. mentioning the row name as well as variable name/column header) in the metadata file. This will help in identifying the notes written for each of the special characters removed with the particular cells during later checking or reviewing of the dataset.

7. Save as CSV file.

8. There should be no space or special character in file name (except "_", "-", ".").

9. File Name, Dataset Title and Dataset Note should come in metadata file.

10. Split sheets (in workbook) as a separate file with sheet names.

Other Things to take care

1. Change format of Data in Data Matrix to Number (remove " , " comma from numbers, if any).

2. Select and delete all blank rows and columns other than the Data Matrix rows and columns.

3. Check Row/column Header for split line (alt+enter which creates space and show " " in notepad) by seeing in the formula bar in expanded state. Entire row or column header should be shown as a running line in the formula bar without showing a new line.

4. While replacing blank cells with NA, select only the data matrix.

5. Remove " , " comma or if required Replace " , " comma with " ; " semi-colon.

6. Variable Name should not be duplicate i.e. each variable name (column header) should be unique. 7. While concatenating; use the " – " (space hyphen space) format to concatenate between text strings.

8. Time Period of Dataset:

a. Please mention the year of the dataset at the end of the title of the dataset (say for dataset named XYZ) e.g. XYZ - 2015 (- YYYY) in case of Calendar Year or XYZ - 2015-16 (- YYYY-YY) in case of Financial Year. It will make dataset unique and avoid the conflict of dataset name when the same dataset for next year or time period is added

Scanned by CamScanner

in data.gov.in. If the dataset mentions a date e.g. 31st December or 31st March, mention it in the Note of the dataset.

b. If the dataset covers a period of time say 2001 to 2015 or 2000-01 to 2014-15, mention time period at the end of the title of the dataset e.g. XYZ from 2001 to 2015 (from YYYY to YYYY) in case of Calendar Years or XYZ from 2000-01 to 2014-15 (from YYYY-YY to YYYY-YY) in case of Financial Years.

c. If the dataset is for/upto a specific or particular date say 20th October 2015 or 11th February 2016, mention the date directly at the end of the title of the dataset e.g. XYZ as on 30-10-2015 (as on DD-MM-YYYY) or XYZ as on 11-02-2016.

d. If the dataset covers a period of time say 2001 to 2015 or 2000-01 to 2014-15, but data for a particular year during this entire period belongs to another date different from the usual date say 31st December in case of Calendar Years or 31st March in case of Financial Years; mention the specific date in case of that particular year in the Note e.g. figures for the year 2015 are as on 30-11-2015 or figures for the year 2015-16 are as on 15-02- 2016. Or within bracket with the year in the dataset itself.

9. Please check the regular expression by opening the file in notepad++ (can be downloaded from net) [^A-Za-z0-9, "'./?\\:&_ \r\n\t()@$#*^%+=<>\[\]!-] Special Note –

1. After preparing the dataset as per above guidelines, Please go through the dataset once again and compare with the original file for reviewing any deficiencies left and correct as required for enhanced accuracy.

2. If you are facing any problem/ ambiguity in following the above mentioned guidelines, Please discuss.

SOP for Quality Assessment of Data

1. Data Compositeness/Completeness/Consistency
2. Check for the constituent elements (variables) within the dataset
3. The dataset should be well explained in terms of the variable present therein the dataset through a descriptive metadata
4. The metadata should well describe the time-period, units, definitions, frequency, data source, jurisdiction and notes to special mention in the dataset
5. The time series data should be continuous in nature
6. Data Coverage
7. Dataset should be made available at the lowest possible levels to allow users correctly describe the phenomena being measured
8. Standard process of "data cleansing"

Scanned by CamScanner

9. Dataset column headers should be self-explanatory and has to be as first row only
10. Assigning string, date, character and numbers to the required fields
11. Abbreviations and acronyms to be replaced by full forms.
12. No special characters and blank spaces (replaced with NA) in the matrix.
13. Dataset should be de-normalized without any merged column
14. No formula of calculated column should appear in dataset like Total or Average of available column or rows
15. Above all it must be in machine readable format viz. CSV, XML, JSON etc. If datasets are released as static CSV while uploading the file, it would convert those datasets to other open formats automatically
16. File name should not contain special character except _ and -; no blank space should not be present in file name.
17. Regular Expression for valid Character for header [^A-Za-z0-9, "'./?;:@_ ()@$#*^%+=\[\]!-]

SOP for publishing as Open Data Norms

Open Government Data

"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 (National Data Sharing and Accessibility Policy0) mandates government departments to proactively open up data. NDSAP in India is applicable to all entities of Government Setup. Identification of Resources (Datasets/Apps) and their organization under Catalogs As per policy each department has to prepare it's 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

Scanned by CamScanner

to be prioritized into high value datasets and non-high value datasets. As per the NDSAP, within a year all the datasets need to be published on the OGD Platform (data.gov.in) and within the first three months at least 5 high value datasets need to be published. Government data generated through following processes and events:

1. Primary Data e.g. Population Census, Education Census, Economic Survey, etc.

2. Processed/Value Added Data e.g. Budget, Planning, etc.

3. Data Generated through delivery of Government Services e.g. Income Tax Collection, MNREGA wage distribution etc.

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

Data Formats

NDSAP recommends that data has to be published in open format. It should be machine readable. Though there are many formats suitable to different category of data. Based on current analysis of data formats prevalent in Government it is proposed that data should be published in any of the following formats:

♣ CSV (Comma separated Values)

Scanned by CamScanner

- XLS (spread sheet- Excel)
- ODS (Open Document Formats for Spreadsheet)
- XML (Extensive Markup Language)
- RDF (Resources Description Framework)
- KML (Keyhole Markup Language used for Maps)
- GML (Geography Markup Language)
- RSS/ATOM (Fast changing data e.g. hourly/daily)

Maintenance And Support

Maintenance of Data Sets/ Feeds: CDO will ensure that published data sets and feeds are up to date and relevant.

Support: CDO will provide required technical and non-technical support over the queries/inputs/ suggestion received from users through email, portal or social media platforms like Facebook, Twitter.

Archiving: CDO will define and set up the process for archiving process. Every data set/feeds catalogue must contain archiving information. Data Sets published over open data portal will be retained as per retention policy. For specific file type (geospatial files), a recent copy must be made available to users through Data Platform.

Ownership: All datasets/ feeds remain the property of publisher, i.e. CDO. The CDO will endorse

Government Open Data License to ensure that published data is not misused or misinterpreted by its users. For more details on Open Data License, please refer: https://data.gov.in/ • Terms of Use: Smart City will publish Terms of Use to restrict the misuse of data and indemnify the city administration in case of any misuse by the end user.

Publishing & Management of Resources (Datasets/Apps)

Contribution of datasets/apps is by login into a simple web-based Dataset Management System through http://www.smartcities.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.

Scanned by CamScanner

Chief Data Officer 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 Chief Data Officer. However, the responsibility on the relevancy and quality of datasets published on the OGD Platform rests with Chief Data Officer.

Metadata Elements for Catalogs/Resources and their Description

Catalog Title (Required): A unique name for the catalog (group of resources) viz. Current Population Survey <Year>, Consumer Price Index <Year>, Variety-wise Daily Market Prices Data, State-wise Construction of Deep Tubewells over the years, etc.

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

Keywords (Required): It is a list of terms, separated by commas, describing and indicating at the content of the catalog. Example: rainfall, weather, monthly statistics. Group Name: This is an optional field to provide a Group Name to multiple 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 applys(ies) to your catalog.

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

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 <Month/Year> etc.

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

Conclusion

Smart Cities Mission intends to implement Data Smart Cities Strategy for all 100 Smart Cities through their ULBs to leverage the potential of data for solving complex urban problems impacting their citizens by encouraging the sharing of non-sensitive data on a common portal (Open Data initiative).

Scanned by CamScanner

Relevant departments and agencies in the respective States will play an important role in helping cities achieve this integration. Two aspects of integration where the State agencies can play a direct role are sectoral data integration in verticals and horizontal integration of data through State data agencies.

30/06/19.

****

Scanned by CamScanner