FINAL
27-01-2021

DATA SHARING AND
ACCESSIBILITY POLICY

Chief Data Officer
Chandigarh Smart City Limited
Chandigarh

TABLE OF CONTENTS

1. **INTRODUCTION**

   Several Public Departments and other organizations in the country as a part of their daily activities generate a large amount of data which can be used by citizens and various stakeholders for economic, scientific and developmental purposes. The Chandigarh Smart city Limited (hereinafter referred to as "The Company") strongly supports the view that the publicly funded data or the data collected through public investment is a public good which must be made available to the public. The Company recognizes the fact that consistent access to data is an essential component of an open, transparent, collaborative and effective government. Therefore it has been decided to implement a "City Data Policy" for the effective communication and coordination between citizens and government by allowing appropriate and responsible sharing of data generated by various public departments, under Chandigarh Administration.

2. **INTERPRETATIONS**

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 collected together for reference or analysis.

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 **Data set** – A named collection of related sets of information composed of separate elements, but which can be manipulated as a unit.

2.5 **Geospatial Data** – All data which is geographically referenced.

2.6 **Information**- Processed data is referred to as information.

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**- It includes non-sharable data as declared by the departments/organizations.

2.9 **Non-Shareable Data**-The dataset which is highly confidential and sensitive in nature is known as non-sharable data

2.10 **Open Access** – Access to data generated from public funding should be easy, timely, user-friendly and web-based without any process of registration/authorization.

2.11 **Open Data** – A dataset is said to be open if anyone is free to use, reuse, and redistribute it.

2.12 **Open List**- It includes sharable and non-sensitive data having open access without any registration or authorization.

2.13 **Restricted Data**- Data which is accessible only through a prescribed process of registration and authorization by respective departments/organizations since it could lead to a threat to life or loss of public assets or critical infrastructure.

2.14 Sensitive Data- Sensitive data as defined in various acts and rules of the Government of India.

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

2.16 Standards – Any application that embeds data handling functions (eg. Data collection, management, transfer, integration, publication etc.).

3. NEED FOR THE POLICY

A large quantum of data which is generated through public funds by various departments, organizations and institutions in the country remains inaccessible to public, although most of such data may be non-sensitive in nature and can be used by public for developmental purposes. Hence it is necessary to formulate a City Data Policy (CDP) which may provide the enabling provision and platform for proactive and open access to the data generated through public funds. Formulation of said policy is also an essence in view of the statutory provision contained in **Section 4(2) of Right to Information Act, 2005**, which is as under:-

"It shall be a constant endeavour of every public authority to take steps in accordance with the requirements of clause (b) of sub-section (1) to provide as much information suo-motu 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"

4. BENEFITS OF POLICY

Easy access to government owned data will enable more extensive use of a valuable public resource for the benefit of the community. By adopting common standards for the collection and transfer of data, integration of individual data sets may be feasible. By sharing data, the need for separate bodies to collect the same data will be avoided resulting in significant cost savings in data collection. 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.

5. OBJECTIVES AND PURPOSES

This Policy is designed to apply to all such sharable data available with the departments either in digital or analog forms. The objective of this Policy is to facilitate access to Company's owned sharable data and information in both human readable and machine readable forms 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 use of public data and information.

6. SCOPE

6.1 This Policy is completely drafted in compliance of National Data Sharing and Accessibility Policy-2012 (NDSAP) which mandates government departments to share non-sensitive/non-confidential data in an open format. The Company in its day to day functioning collects and generates

large amount of such non-sensitive/non-confidential data from several Govt. departments/organizations in Chandigarh which generally remains inaccessible to the public. Such data need to be available in an open format, meaning thereby, it should be free from any license, registration or any other mechanism of control. Data in such an open format would certainly enhance transparency and accountability while encouraging public engagement.

6.2 The Company undertakes to organize workshops and other training programmes that would engage the stakeholders and researchers (eg. Academia, students, research institutes, startups etc.) to understand the processes involved in Contribution, approval, identification, and assessing data needs. Such events and programmes so conducted by the Company will enhance the intentions among the stakeholders and researchers to promote a technology-based culture of data management and it will also led to positive citizen outcomes. The training so imparted by the Company will certainly help to engage stakeholders to solve the major urban challenges by assessing the data need.

6.3 This Policy will apply to all the data and information created, generated, collected from other Govt. Departments/Municipal Corporation directly or indirectly and archived by Chandigarh Smart City Limited. This policy applies to any person/user, organisation, administrators, contractors, etc. who intends to access information or assets through data portal of Chandigarh Smart City Limited. Specifically, the Data Policy applies to the following information:

- Data/Information collected, captured, aggregated, processed and shared by various departments.
- Citizens data/Information

7. DATA MANAGEMENT AND TEAM STRUCTURE

A data team be set up for management of data at city level for enabling coordination between various departments for making sure that data is ready and available for access as and when required. The quality of data and its use largely depends on the data team. Key officials of the data team and their roles and key responsibilities are defined as under:

7.1 CITY/CHIEF DATA OFFICER (CDO)

The CDO will be the officer responsible for implementation of data smart cities strategy at the city level. CDO will work with city leadership to assess and tap the potential of data and set up data culture across the organisation and outside the organisation. The key responsibilities of CDO are as under:

- To ensure that published datasets are without any ambiguity and same are as per the needs of various users in the city.
- To coordinate with Mission Director Officer (MDO) to align with Mission Data strategy and priorities with respect to open government initiatives and policies.
- To organize regular meetings of the City Data Alliance (CDA).

- To coordinate with officers of various other government departments/organisations within the city for effective implementation of Data Strategies.
- To coordinate with Data team to assess the data requirement of various stakeholders in City ecosystem.
- To publish data catalogues and datasets on Open Government Data Portal and ensure that such datasets are updated at regular intervals.
- To identify the high value datasets and schedule their release and publication on Open Government Data Platform.

7.2 DATA COORDINATORS

Data Coordinators will assist CDO at the department/government agency level as reporting staff. The key responsibilities of Data Coordinators are as under:

- Aggregate the data demand from various channels.
- Sensitizing the department employees over the importance of data quality etc.
- Perform collection, interpretation and recording of data in accordance with CDP standards
- Perform data validation and ensure data quality.
- Sort and organize the data; both hard copy and electronic versions.
- Assist department staff in data entry when required.
- Provide data management updates in all internal and external meetings as required.
- Analyse data for quality improvement purposes.
- Prepare and submit data required for audits.
- Transmit data report to City Data Office.

8. DATA MANAGEMENT PRINCIPLES

8.1 Data Categorization

Data can be categorized into two broad categories which is as under:-

- **Personal Data:** - Personal data is that data which is specific to a particular individual which is confidential in nature and it is the responsibility of civic administration that such personal identifiable data/information shall not be furnished on open data platform.
- **Non-Personal Data:** - Non-Personal data is that data which is not specific to a particular individual and which cannot be referenced to any particular individual. Such data shall be treated as non-confidential in nature which can be accessed by the public.

8.2 HOW DATA SETS ARE IDENTIFIED/CATEGORIZED

The company shall identify data sets by the following categories: -

Negative List

The datasets which are confidential and personal in nature and datasets which are in the interest of nation's security and are non-shareable in nature shall not be open to public and such category of data shall be

placed in Negative List which shall include the non-shareable and highly confidential data.

Open List

All other datasets which do not fall under the category of negative list would be covered in the Open List that includes the data which are sharable, public, no--personal and non-sensitive in nature and accessibility of such data should be in an open format.

Restricted List

Datasets which are declared restricted by the Government of India, or by any statute/act or by any policy/notification issued by government, would be covered under the Restricted List. Accessibility to such data is generally restricted but can be permissible by following the prescribed process of authorization.

8.3 CLASSIFICATION OF DATA

Different types of data sets generated in any form can be classified as sharable data and non-sharable data.

Sharable Data

The data sets which are non-confidential and non-sensitive in nature are known as sharable data. Such data should be accessible in an open format.

Non-Sharable Data

The data sets which are highly confidential in nature are known as non-sharable data. Such data cannot be shared to the public.

Restricted Data

The Data sets which are accessible only through a prescribed process of registration and authorization by respective departments and organisations.

8.4 TYPES OF ACCESS

Open Access

Datasets which are sharable and non-sensitive in nature and which are included in the open list are easily available through open access without any process of registration and authorization.

Registered Access

Datasets which are non-sharable and Confidential in nature and which are included in negative list are accessible only through a prescribed process of registration/authorization. This process is to ensure that the Company can track the users and the use of data. The users are required to register their names through the web portal and then download the information by using the user name and password provided to them at the time of registration.

Restricted Access

Datasets which are declared restricted by the Government of India, or by any statute/act or by any policy/notification issued by government, are accessible only through and under authorization.

8.5 DATA FLOW/APPROVAL FRAMEWORK

Once the data is collected, various stakeholders may be involved in its movement i.e. from collection to publication after approval of competent authority. It would be the responsibility of the data team to maintain the authenticity and correctness of the data so collected and published thereafter. There must be a specific process to manage the flow of available data and a formal framework that defines the movement of data among various stakeholders. The concerned stakeholders shall approve and authorise the data usage.

The approval framework for Chandigarh Smart City Ltd. is as under:-

Steps	Process	Responsibility
I	Collection of data	The concerned employees will collect the data and will submit it to the data coordinator of the department
II	Verification of data	Data is to be verified by the data coordinators and data champions
III	Data Validation	City Data officer
IV	Data Approval	City Data Officer/ Municipal Commissioner

8.6 DATA ARCHIVAL AND RETENTION

Data Archival

As per National Data Sharing Accessibility Policy-2012 (NDSAP-2012) Data Archive is defined as a place where machine-readable data are acquired, manipulated, documented, and distributed to others for further analysis and consumption. In layman language it may be defined as a collection of historical records which is still important and necessary for future reference. Question here arises that when a data can be considered as a historical data and for knowing same there is need of a specific time frame during which data can be stated to be in retention stage and after that specific time data can be move to the archive stage. Therefore, data older than 6 months may be considered as historical data and may be archived for further analysis and consumption.

Data Retention

Retention may be defined as the maintenance and upkeep of data which can be accessed by an authorized user as and when required. It is the CDO to undertake retention of data i.e. maintenance of data and retrieval schedule of the data will be as per the rules and regulations defined by the

Government of India. Datasets collected from various departments/organisations must be retained in a live state for a certain period before its archival. Different ways and place for retentions of data depends upon the category of data i.e. for example the data which is in electronic form may be stored on shared drives and the data which is in physical form may be kept in files or any other storage units assigned to each department.

8.7 DATA SECURITY AND PRIVACY

This Policy encompasses the systems and processes that ensure not only the integrity but also the security of datasets published by various contributors or departments. The datasets which are highly confidential and sensitive in nature and involves the question of national security and privacy of an individual must be certainly secured and access to such datasets should be restricted as a general principle and if made permissible that should only be through a prescribed process of registration or authorization. It has to be ensured that privacy of personal data and confidentiality of sensitive data is protected from any loss, unauthorised use and corruption. All data flows, storage and sharing should be adhere to the national and state level data privacy and security policies already in place. Data dissemination should be only to authenticated and authorized stakeholders.

9. STANDARD OPERATING PROCEDURES (SOPs)

9.1 SOP for Data Collection

Data Collection is the first step towards Data driven decision making and evidence based governance. Resources (Datasets) are to be contributed by various departments in a predefined workflow i.e. after taking approval from the concerned Authority. Data Contributors nominated by the departments are authorized to publish datasets in an open format on the web-portal of the Department. Data is to be identified and prioritize for release of datasets and categorize the type of access granted for them and publish as many high value datasets as possible. Each department shall have its own criterion of high value and low value datasets. However High value Data is governed by the following principles which shall be taken care of by the departments while publishing them.

- Completeness
- Primary
- Timeliness
- Ease of Physical and Electronic Access
- Machine Readability
- Non-discrimination
- Licensing
- Permanence

9.2 SOP for Electronic Data Collection

Datasets released by several departments should be accessible with ease, either through physical or through electronic means. **Physical Means** include requirements to visit in a particular office in person or

requirements to comply with particular procedure (such as completing forms or submitting requests and representations etc.) Automated **Electronic means** include datasets accessible only via submitted forms or systems that require browser-oriented technologies (e.g. Flash, Javascript or java applets, mail etc.)

- While collecting such data, the IT applications/IT systems should be developed in such a way that they follow proper checks and conditions, and under any circumstances do not accept null data which is of no use.

- Special care need to be taken for privacy and security of such data. The measures mentioned in the above sections of the policy shall be followed to maintain confidentiality and security of such data.

- It is recommended 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:

  ➢ CSV (Comma separated values)

  ➢ XLS (Spread sheet - Excel)

  ➢ ODS (Open Document Formats for Spreadsheets)

  ➢ XML (Extensive Markup Language)

  ➢ KML (Keyhole Markup Language used for Maps)

  ➢ GML (Geography Markup Language)

9.3 SOP FOR DATA PROCESSING AND CLEANING

Raw data may be old and inaccurate and can have an adverse impact on results. In order to make data more meaningful and useful, it must be processed. However before processing the data, the datasets must be claimed by making necessary corrections as required. Data cleaning is needed to ensure that data is correct and consistent and is free from any kind of errors or discrepancies. Following steps need to be followed for purpose of data processing and cleaning:

- Collection of data from departments/organisations
- Inspection & Analysis, ensuring accuracy of collected data
- Identification of irrelevant data.
- Identification of errors, inconsistencies in data
- Standardisation of datasets after appropriate diagnosis
- Editing of data by making corrections, deletion, removing extra space
- Finalization of datasets
- Publication of corrected datasets after verification.

9.4 SOP FOR QUALITY ASSESSMENT OF DATASETS

Quality Assessment of data is needed to ensure that the quality standards of data are duly met with. Data must be accurate, free from any sort of legal issues, privacy of an individual is maintained and does not compromise with the National security. Data quality shall be assessed

from the perspectives of adequacy, appropriateness, accuracy and reliability, authenticity, consistency and validity.

- Responsibility for quality check of datasets rest with the CDO while publishing data on the open data portal or sharing it with stakeholders upon request.
- Under some circumstances, a special committee under the leadership of Worthy Commissioner, Municipal Corporation, Chandigarh, comprising of CDO, and Data coordinators may be formed for data quality assessment.

9.5 SOP FOR DATA PUBLISHING

The standards for publishing datasets and information on the Open Data Portal are very well defined in National Data Sharing Accessibility Policy-2012 (NDSAP-2012) which must be taken into consideration while publication of data. It is the responsibility of City Data Officer that data should be published as per the open data norms and in accordance with the standards as defined in NDSAP-2012. Following steps are required to be followed for publication of data on Open Data Portal:

- Proper procedure for collection of relevant and valid data must be followed. The collected data must not contain personalized information and it should be authenticated and free from defects.
- It must be ensured that the datasets must be published in an open Format and should be machine readable. Data Format can be chosen from the list recommended by NDSAP.
- The selection of datasets and data publishing options should be carefully studied before publishing the data on Open Data Portal.
- It should also be ensured that all the errors and inconsistencies are removed from the primary data and that the dataset included all the metadata alongwith referential information.

9.6 SOP FOR ENGAGING STAKEHOLDERS

The concept recognizes the value of enhancing engagement among all four stakeholders of the quadruple-helix model i.e. Government, Citizens, Academia and Industry. City Data Officer alongwith Data team/ Data Coordinators must assess the data requirements of various stakeholders in city. Key activities may include:

- Identifying stakeholders from various age groups and ethnicities and engaging them in city initiatives.
- Organising workshops, hackathons/events to promote brainstorming over requires datasets.
- Decision making and consultation with data experts to zero down on the most important datasets required on the portal.
- Data ideation with public forum to gain understanding of citizen/industry needs.
- Formation of city alliances.

9.7 SOP FOR DATA COLLECTION, PROCESSEING AND ANALYSIS FOR ONFIELD SURVEYS

Field surveys may be defined as collection and gathering of data and information at the local level by conducting primary surveys. In case of inadequate and inaccurate existing datasets, o-field survey plays a significant role in fulfilling the certain gap by removing the inadequacy and other defects from the existing datasets. Survey data processing consist of four important steps:

- Survey designing
- Data collection
- Data processing
- Data analysis

9.8 SOP FOR DATA ANALYSIS

Data analysis is a multi-dimensional field that uses mathematical, statistics, predictive modelling and machine learning techniques to find meaningful patterns and knowledge in recorded data. Apart from presenting the actual data, cleaned datasets might also be used to discover new patterns or analyse existing patterns, trends or behaviours. Data analysis may help multi-disciplinary researchers provide different perspectives or even solutions on civic issues like transport, traffic, solid waste etc.

10. TECHNOLOGY FOR SHARING AND ACCESS

This Policy is to support the Government of India in its commitment to transparency and efficiency in governance. Therefore, the data is to be made available through a web-based portal that is clear and simple for everybody to use. The shared data should be accessible to all users irrespective of device in use, technology or ability. The main aim behind web-portal would be to provide maximum accessibility and usability to its visitors.

The data repository will hold current and previous data. The data is to be updated periodically from time to time. Online analytical processing of data will be available on the web portal. The main features of the web portal will include:

- Secure Web Access
- Searched based Report
- Dynamic report in exportable format
- Complete Metadata
- Bulletin Board

11. CITIZEN ENGAGEMENT AND SUGGESTION

The Open data Platform shall also have a strong component of Citizen Engagement. Citizens can express their views and suggestions and as well rate the datasets published by the department. Datasets can be rated by public on three principal aspects i.e. Quality, Accessibility and Usability. The users can also embed the resources/datasets in their blogs or websites. Facility to contact the concerned authority shall also be available on the Platform.

12. CITY DATA ALLIANCE

The Policy not only supports the idea of sharing and accessibility of data to the public, but it also promotes the idea for creation of an alliance or association which will provide a collaborative framework to solve critical city problems through use of data. The alliance will undertake education and awareness about data in the community, understand and address concerns on data privacy and security. It shall further aim to create data collaborations between various government and private agencies for solving various problems relating to data sharing and continuously evolve the culture of data sharing in the city.

13. BUDGETARY SYSTEM

The implementation of this Policy is expected to entail expenditure for both the Company as data owner and for data managers also for managing data refinement, data storage, data security, data quality and data upgradation etc. Therefore, budgetary provisions and appropriate support for data management to the Company by Government of India would be necessary. It is for the Government to design and position a suitable budgetary incentive system for the data owner for enhancing and encouraging open access to the sharable data.

14. CONTACT US

We seek from the users to inform the company of any analysis or publications resulting from their work with the data set. This will help us to maintain and overview of how data is being used. We are very interested to learn of any secondary analysis of our data and about forthcoming publications that may arise from work with our data. Please contact us: -

- If you would like to be kept informed of future data workshops.
- For any further information about our data or our methodology.
- If you would like to have access to more data

15. CONCLUSION

The policy provides for the facility of optimum accessibility and usability of data by the users and further encourages the proper organization of data, with better access services. Data Sharing and Accessibility Policy aims at the promotion of a technology – based culture of data management as well as data sharing and access. It is necessary to open up the information on the available data which can be shared with civil society for developmental purposes.