City Data Policy

for

Surat Municipal Corporation

Chief Executive Officer
Surat Smart City Development Ltd.

Message from the Desk of Municipal Commissioner, Surat Municipal Corporation

With the emergence of Smart Cities and data-driven innovation hubs across the country, the potential of data has become undisputed. Recognizing that timely and consistent access to data is an essential component of an open, transparent, collaborative and effective government, the city has decided to implement a City Data Policy (CDP) for the effective communication and coordination between citizens and government.

The main aim for designing this policy is to ensure data and information is utilized to its maximum potential. This policy will provide a framework for easier and effective data sharing among stakeholders and it is within the legal and legislative framework defined by the Surat city. It is expected to trigger a higher quotient of trust towards governments, increase collaboration and engagement with citizens, public and private entities, lead to innovation-driven problem-solving, and eventually a higher quality of life for citizens. For this to happen, it is of paramount importance to have a set of guidelines with necessary conditions built in to avoid misuse, while also generating awareness among various stakeholders and the steps and processes that need to be put in place to enhance impact of data.

For the effective and timely implementation of this policy, I urge the city and all the stakeholders to treat this policy document as a call for participation in our joint effort towards creating a robust open governance and innovation ecosystem for a more livable urban future.

(--sd--)
Shri Banchhanidhi Pani
Municipal Commissioner,
Surat Municipal Corporation

Chief Executive Officer
Surat Smart City Development Ltd.

Abbreviations/Definitions

- **Data:** Data refers to 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.
- **Data Archive:** A place where machine-readable data are acquired, manipulated, documented and distributed to others for further analysis and consumptions.
- **Data generation:** Initial generation/collection of data or subsequent addition of data to the same specification.
- **Dataset:** A named collection of related sets of information composed of separate elements, but which can be manipulated as a unit.
- **Geospatial Data:** All data which is geographically referenced.
- **Information:** Processed data is referred to as Information.
- **Metadata:** Metadata is data about data. The information that describes the data source, and the time, place, and conditions under which the data were created. Metadata informs the users of who, when, what and where data were generated. Metadata allows the data to be traced to a known origin and known quality.
- **Negative list:** List of prohibitive datasets/feeds, deemed non-shareable by the departments/ organizations.
- **Restricted Data:** Data which are accessible only through a prescribed process of registrations and authorization by respective departments/organization since it could lead to a threat to life or loss of public assets or critical infrastructure.
- **Shareable Data:** The data not covered under the scope of negative list and non-sensitive in nature falls under shareable data.
- **Standards:** Any application that embeds data handling functions (e.g. data collection, management, transfer, integration, publication etc.).
- **Open Access:** Access to data generated from public funding should be easy, timely, user-friendly and web-based without any process of registration/authorization.

Chief Executive Officer
Surat Smart City Development Ltd.

Contents

1. Need for Data Policy 5
2. Scope of the Data Policy 6
3. Stakeholders and Collaboration 7
4. Data Management and team Structure 9
5. Roles and Responsibilities 10
    1. City Data Officer (CDO) 10
    2. Data Champions (DCS) 11
    3. Data Coordinators 11
6. Data Management Principles 12
    1. Data Categorization 12
    2. Data Classification 12
    3. Data Flow/Approval Framework 13
    4. Data Archival and Retention 13
    5. Data Security and Privacy 14
7. Standard Operating Procedures (SOPs) 15
    a) SOP for Data Collection 15
    b) SOP for Electronic Data Collection 15
    c) SOP for Data Processing and Cleaning 16
    d) SOP for Quality Assessment of Datasets 17
    e) SOP for Data Publishing 17
    f) SOP for Engaging Stakeholders 18
    g) SOP for Data Collection, Processing and Analysis for On-Field Surveys 18
    h) SOP for Data Analysis 19
8. Annexure 20

Surat Municipal Corporation

Chief Executive Officer
Surat Smart City Development Ltd.

1. Need for Data Policy

City Government departments generate a large amount of data. The data is generated as part of daily activities of the municipal corporations and city governance departments. This raw data, if used appropriately, can help generate valuable information for Surat Municipal Corporation as well as entities external to Surat Municipal Corporation. This data can be used by these various stakeholders for economic, scientific and developmental purposes.

The need to facilitate sharing and utilization of this large amount of data primarily points to the need of a structure defining the rules and regulations. Today most of the government information is stored in databases and is not effectively utilized for public good. The current regime of data management does not enable open sharing of Government-owned data with other arms of the Government, nor does it expect proactive disclosure of shareable data available with data owners. Such regimes could lead to duplication of efforts and loss of efficiency in planning activities focused on city development. Hence, City Data Policy (CDP) of Surat Municipal Corporation aims to provide an enabling ecosystem and a platform for providing proactive and open access to the data generated through public investments and public revenue available with various departments of Surat Municipal Corporation/Surat Smart City Development Limited, other government departments etc. This document is intended to put in place a formal data governance mechanism at Surat Municipal Corporation and is an attempt to further enhance the existing data initiatives of Surat Municipal Corporation by having in place a robust, complete and inclusive CDP. Having such a policy in place will make the city of Surat being identified as one of those cities that is data-driven and data-sufficient.

Chief Executive Officer
Surat Smart City Development Ltd.

2. Scope of the Data Policy

This Policy document covers principle considerations concerning the use of data (access and utilization of datasets including appropriate privacy management), and the principles governing the data sharing program for Surat; thereby defining the expectations for departmental participation and governance of the data program.

The policy is intended as a resource for city administrators such as Municipal Commissioner, Smart City CEO, and other officials such as City Data Officer, heads of various government departments, Data Coordinators, Data Champions and external agencies - parastatal, civic, private) interested in engaging with the data initiatives of the City.

This Policy will apply to all data and information created, generated, collected and archived by Surat Municipal Corporation/Surat Smart City Development Limited. This policy applies to any person/user, organization, administrators, contractors, etc. who intends to access information or assets through any data portal of Surat Municipal Corporation/Surat Smart City Development Limited. Specifically, the Data Policy applies to the following information assets of Surat Municipal Corporation/Surat Smart City Development Limited:

1. Data/information collected, captured, aggregated, processed and shared by Surat Municipal Corporation/Surat Smart City Development Limited
2. Citizens data/information
3. Personnel data/information relating to employees of Surat Municipal Corporation/Surat Smart City Development Limited

Chief Executive Officer
Surat Smart City Development Ltd.

3. Stakeholders and Collaboration

To assess, strategize, plan, implement and review the City Data Policy, SMC/SSCL will form a City Data Alliance which may comprise the members from the following:

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

b) **Funding agencies:** Funding agencies which regularly work with City Administration in different domains for e.g. World Bank, ADB, DFID, etc.

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

d) **Academia:** Representatives from leading universities/colleges/schools in the 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 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:** 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 services like Doctors, CA, and Engineers etc.

City Data Alliance (CDA)

The CDA will provide a collaborative framework to create and define use cases to solve critical city problems through the use of data, catalyze the right set of collaborations and networks to make available such data and undertake continuous dialogue between various stakeholders in the city around the CDP so as to inform and evolve the CDP effectively.

The responsibilities of the CDA may include:

a) To act as an advisory group to the city leadership on the CDP.

b) To assess the data needs of various smart city stakeholders.

Chief Executive Officer
Surat Smart City Development Ltd.

c) To promote data driven governance and policy formulation.

d) To design and implement solutions based on city data.

e) To support industry to design solutions using emerging technologies like Artificial Intelligence (AI), Machine Learning (ML) and Blockchain.

f) To assess and design use cases critical to the citizens of the city.

g) To generate awareness among various stakeholders towards open government initiatives.

h) To bring city's stakeholders on a 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 the city.

k) To deliver research papers using city data on civic problems in the city.

l) To design and develop prototypes/ solutions on civic problems in the city.

m) To organize data-challenges on complex civic problems.

n) To organize hackathons and support shortlisted solutions at city level.

o) To set up internship/training program for postgraduate and graduate students to work with the City Data Team along with the CDO.

p) To publish the periodic progress report.

q) To prioritize the datasets/feeds for publishing on the open data platform.

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

s) To support, engage and encourage network/groups/members of data enthusiasts in the city.

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

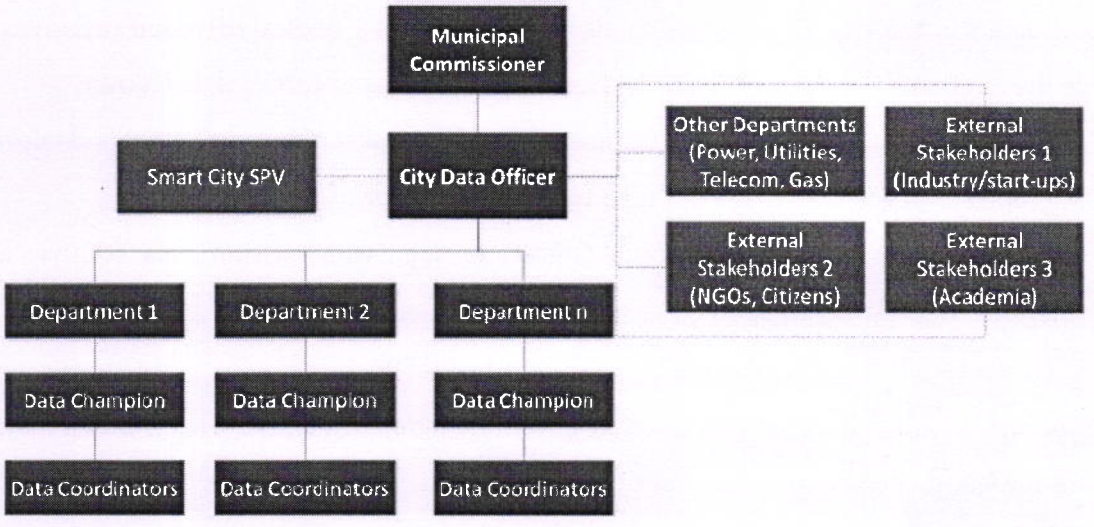
u) To support CDOs by extending resources (like interns, researchers, technology experts), funds (program sponsorship etc.) and technology (solutions etc.).

v) To share data available with the partners on data platforms to promote city data.

Chief Executive Officer
Surat Smart City Development Ltd.

4. Data Management and team Structure

Surat Municipal Corporation has set up a data team for management of data at the city level and enabling coordination between various departments for making sure that the data is available as and when required. The quality of data available and its use depends largely on the efficacy of the team that is put in place. The data team structure for Surat Municipal Corporation is as follows:



Surat Municipal Corporation will also engage and secure buy-in from both internal and external stakeholders on key decisions. Further, this will help navigate through complicated hurdles (e.g. bureaucratic, political etc.) and to take prompt decisions and actions pertaining to collection, segregation and release of data.

Chief Executive Officer
Surat Smart City Development Ltd.

5. Roles and Responsibilities

The activities under this will be carried out through key officials of the data team, whose roles and key responsibilities are defined below:

1. City Data Officer (CDO)

The CDO will be the officer responsible for implementation of the DataSmart 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. CDO will report directly to City Leadership and act as single point of contact to all internal and external stakeholders in the city. City leadership also needs to deploy dedicated skilled resources to drive the data initiative through CDO. The key responsibilities of CDO are as follows:

a. Ensure that the CDP 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 the City Data Alliance (CDA).

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

e. CDO along with team of Data Champions/ Coordinators must assess the data requirements of various stakeholders in smart city ecosystem. External stakeholders may also need to be engaged to understand the data needs. CDO must engage various internal stakeholder at operational, tactical and strategic level to assess the data need to make decisions. Data needs and frequency of consumption needs to be outlined for internal stakeholders.

f. Publish Data Catalogues and Datasets/Feeds on Open Government Data Portal and ensure that such datasets are updated at regular intervals as needed and create mechanisms for continuous feedback from citizens and stakeholders on type of datasets to be published.

g. Assess all the operational IT Projects for identifying public datasets/feeds. Data Champions and Data Coordinators in respective departments must prepare integration plan with respective IT vendor/ integrator to ensure compliance as per CDP.

h. Assess all proposed or under implementation projects to identify the datasets/feeds which could generate public datasets/feeds or may be useful for internal analysis. CDO must work with concerned System Integrator/vendor to ensure compliance of smart solutions with CDP.

Chief Executive Officer
Surat Smart City Development Ltd.

i. Assess all periodic and recurring MIS needs to identify the datasets/feeds which could be shared with other departments through data exchange. CDO will also assess third party funded reports related to city operations for e.g. City Mobility plan, Health Plan etc.

2. Data Champions (DCS)

Data champions will be senior functionaries who would champion the implementation of the CDP in their respective departments/ organisations. Their responsibilities are as follows:

a. Shall identify the datasets/feeds, derived information, intelligence or data challenge with respect to day to day operations of the department.
b. Actively publish/enable publishing of datasets/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 CDP.
c. DCs will be assisted by the Data Coordinators within the department to streamline processes of data reporting, collection and analysis etc. DCs will be responsible for data quality.
d. Undertake activities to engage with their stakeholders and evolve their department's strategy on data in line with the deliberations.

3. Data Coordinators

Data Coordinators will assist Data Champions at the department/government agency level as reporting staff. Their responsibilities are as follows:

a. Aggregate the data demand from various channels.
b. Sensitizing the department employees over the importance of data quality etc.
c. Perform collection, interpretation and recording of data in accordance with CDP standards and CDO guidelines.
d. Perform data validation and ensure data quality.
e. Sort and organize the data; both hard copy and electronic versions.
f. Transmit data report to Surat Municipal Corporation/Surat Smart City Development Limited or CDO via Internet.
g. Update Surat Municipal Corporation/Surat Smart City Development Limited website or Surat Open Data Portal with latest data records.
h. Assist department staff in data entry when required.
i. Provide data management updates in all internal and external meetings as required.
j. Analyze data for quality improvement purposes.
k. Prepare data for reporting, meetings and presentations for the concerned department and Surat Municipal Corporation/Surat Smart City Development Limited at large.
l. Ensure data management procedures comply with CDP.
m. Provide statistical analysis and longitudinal analysis of data.
n. Prepare and submit data required for audits.

Chief Executive Officer
Surat Smart City Development Ltd.

6. Data Management Principles

Implementing a data policy calls for a process that takes care of various aspects of data management. Having such a process in place will impart a sense of uniformity in the way in which data is generated/ created to the point that it is destroyed/refined/reused. The following section defines the broad process that may be implemented for publishing the data under Open Data initiative.

1. Data Categorization

Data will be categorized into two broad categories:

- **Personal Data:** Personal data is that data which is specific to a particular individual. It is the responsibility of the civic administration that they do not, in any case, publish personal identifiable data/information or parts of personal data/information on any of their Open Data Platforms or Datasets. It becomes the responsibility of the CDO to ensure that all personal data is anonymised before it is published.
- **Non-Personal Data:** Non-personal data is that data which cannot be identified or referenced to any individual. Anonymous data is also Non-personal data when all personal indicators and identifiers are eliminated for that particular data element.

2. Data Classification

Data classification is the process of organizing data into categories for its most effective and efficient use. Further to data categorization, there is a need to classify the data basis its intended usage and stakeholders. There may be some data which can be open to the public, while some other may be confidential and restricted. Such distinctions need to be appropriately defined to prevent misuse and maintain confidentiality. Surat Municipal Corporation will prepare the negative list of data which will be periodically reviewed. Further, all datasets will be tagged such as open data, shareable, or others.

Personal and non-personal data will be broadly classified into four levels:

Classification	Class	Definition
Level 1	Public/Shareable Data	Those data not covered under the scope of negative list and nonsensitive in nature. This data is available for public consumption and use.
Level 2	Negative List	Non-shareable data as declared by the departments/organisations.
Level 3	Restricted Data	Data which are accessible only through a prescribed process of registration and authorization by respective departments/ organisations.
Level 4	Sensitive Data	Sensitive data as defined in various Acts and rules of the Government of India.

Chief Executive Officer
Surat Smart City Development Ltd.

3. Data Flow/Approval Framework

Surat Municipal Corporation will set up enterprise processes to control the existing available data within the City administration. At every stage of data generation, the concerned stakeholders shall approve and authorise the data usage. Data flows may vary according to different scenarios, such as data being circulated between departments, uploading data on the open data portal, sharing data with third party, etc. Each dataset has a trustee accountable for data quality and security. Appropriate data flow and approval mechanisms should be in place for such situations. Also, all applications must be developed in compliance with the Government of India's India Enterprise Architecture (INDEA) framework. The Approval framework for Surat city data is as follows:

Process	Data Collection	Data Verification	Data Validation	Data Approval
Responsibility	Data Champion Data Coordinators	Data Champion, City Data Officer	City Data Officer	City Data Officer CEO, SSCDL

4. Data Archival and Retention

Data Retention

Necessary measures to be taken to undertake retention of data, i.e. the maintenance of documents to be accessed by an authorized user at a later stage. Retrieval schedule of the data will be as per the rules and regulations applicable to Surat Municipal Corporation:

1) **Electronic:** Records to be retained in electronic form. These records may be stored on shared drives with access to only authorized individual/ group of individuals.
2) **Physical:** Record to be retained in physical form. These records may be kept in file cabinets or any other storage units assigned to each department with proper labelling so as to enable quick identification of the records.
3) **Original Form:** Records to be retained in the original form in which they were created or used i.e. either electronic or physical.

Data Archival

Necessary measures to be taken to set-up the process for archiving data. Every dataset/feeds catalogue should contain archiving information. For specific file types (e.g. geo-spatial files), recent copy will be made available to users through Data Platform.

E-Files/records may be digitized by any one of the categories:

1) Category-I (e-Files/records to preserved permanently which are of historical importance) – For 10 years, it will be kept in the Department's server and thereafter transferred other available physical storage formats such as Tapes, hard-drives, Storages etc.
2) Category –II (e-Files/records of secondary importance and have a reference value for a limited period) – 10 years on the Department's server. In exceptional cases, if the record is required to be retained beyond 10 years it will be upgraded to Category-I.

Data will be stored in the main database for 6 Months in a live state so that whenever a report needs to be generated, the data will be extracted from main database. Data older than 6

Chief Executive Officer
Surat Smart City Development Ltd.

months will be archived. If report duration extends beyond 6 months, the data will be retrieved from archival to generate the report.

5. Data Security and Privacy

Necessary measures will be taken to ensure that data is protected from loss, unauthorized use and corruption, through adoption of international standards and best practices, duly protecting the privacy of personal data and confidentiality of sensitive data. All data flows, storage and sharing should adhere to the National and State level data privacy and security policies already in place. In case of any grievance, the associated parties may reach out to the CDO/Municipal Commissioner/Smart City CEO for resolution. Security measures needs to be followed during the collection and management of the classified data at all department levels, including:

- Minimize collection of personal data
- Delete data that is no longer necessary
- Restrict access to only those who need it
- Secure data throughout its entire lifecycle

The following security features will be followed while managing data in the city government:

- Data dissemination should be only to authenticated and authorized stakeholders (both internal and external) through data fiduciaries.
- The usage rules for data elements must specify for what purposes the data can or cannot be used. The solution should support SSL encryption mechanism for transferring data across network.
- The data transferred across network should be encrypted using Public Key Infrastructure (PKI).
- Access to all system resources including data files, devices, processes and audit files should be provided to the intended users only.
- All mobile applications should be designed and developed in a way that it ensures security of the application and data on the device.
- Ensure to protect documents by assigning security parameters and criteria in order to provide more effective protection for an electronic document in order to maintain Confidentiality, Authorization, Accountability, Integrity, Authenticity and Non-repudiation.
- Database Activity Monitoring (DAM) should be available to monitor all databases.

Chief Executive Officer
Surat Smart City Development Ltd.

7. Standard Operating Procedures (SOPs)

Standard Operating Procedures (SOPs) are succinct guidelines designed to achieve consistency in specified situations by postulating a standard practice in performing those functions. These SOPs are designed with a view to enhance and standardize data operation and management. The following SOPs may be implemented for various activities under this initiative and will be subject to modifications from time to time as per the requirement.

a) SOP for Data Collection

Data collection is referred to as the method of collecting information in a systematic way. This is the first step towards data-driven decision making and evidence-based governance. Data collection provides both a baseline to measure and a target to improve.

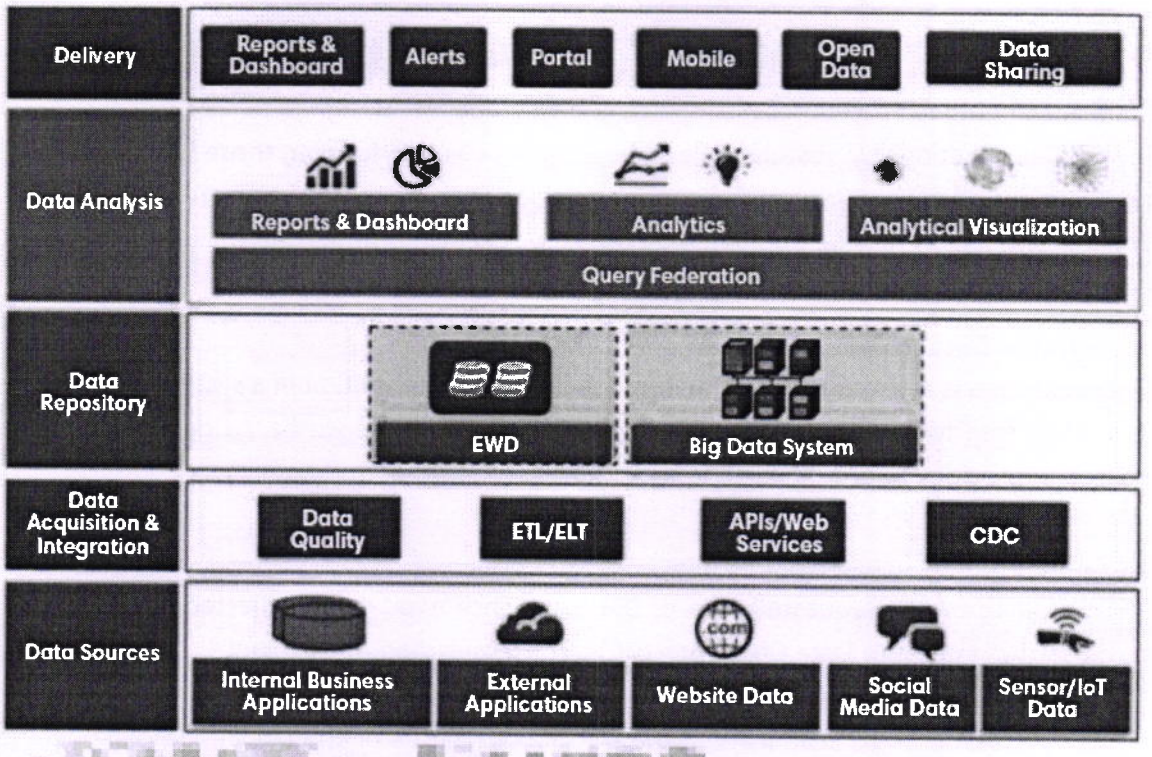
1. Request for any data should be directed to the CDO.
2. Depending on the requested data or the data which need to be collected, the CDO shall direct the request to Data Champion (DC) of the concerned department.
3. The DC of the concerned departments checks for the requested data. If data is available with department, DC shall instruct the Data Coordinators or the concerned personnel to gather the data in requested format.
4. DC will take approval of data from their HOD.
5. If fresh data is needed to be captured/acquired, the Data Champion/Data Coordinators in consultation with CDO shall take appropriate action.

b) SOP for Electronic Data Collection

Additional factors need to be kept in mind when any form of data is collected, accessed, transferred or stored electronically. Cities ecosystem comprises of various technology solutions ranging from Sensors, IoT, SCADA, Electronic camera, GIS, payments system etc. which generate loads of structured data every second on different dimensions. Cities could also leverage various unstructured data emerging from different sources and third-party systems like social media, internet, websites, videos, images etc.

1. Necessary measured to be taken for data privacy and security. The measures mentioned in above sections of the Policy shall be followed to maintain confidentiality and security of data.
2. For data collection and integration with the IoT systems / sensors, the following Data Architecture highlighted below shall be followed by Surat Municipal Corporation/Surat Smart City Development Limited.

Chief Executive Officer
Surat Smart City Development Ltd.



c) SOP for Data Processing and Cleaning

Raw data may be old and inaccurate and can have an adverse impact on results. Data cleaning will be done to ensure that data is correct, consistent and useable by identifying any errors or corruptions in the data, correcting or deleting them, or manually processing them as needed.

1. While collecting the electronic data, the IT applications/ IT systems should be developed in such a way that under any circumstances these applications/systems should not accept any wrong data/null data.
2. If there is existing data, identify discrepancies which may come from different sources.
3. The collected data shall be properly processed and cleaned before performing any kind of analysis.
4. If needed, commercial software available in the market can be used with prior approvals from the concerned authorities.

Extract, Transform and Load (ETL) is the common methodology used for data integration and processing. It is a three-step process which used for data integration to blend data from multiple sources. It's often used to build a data warehouse. During this process, data is taken (extracted) from a source system, converted (transformed) into a format that can be analyzed, and stored (loaded) into a data warehouse or other system.

Specifically, Data Champion/ Data Coordinator will keep in mind the following points while collecting data from their respective department:

- Spaces in extra columns Compliance
- Ward-wise Compliance
- Blank Cells Compliance
- Standard format Compliance

Chief Executive Officer
Surat Smart City Development Ltd.

- All NA Compliance
- Special Characters Compliance
- Split Sheets Compliance
- Datasets Compliance
- Data Completeness

d) SOP for Quality Assessment of Datasets

Quality assessment of data is needed 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. Data quality shall be assessed from the perspectives of adequacy, appropriateness, accuracy and reliability, authenticity, consistency and validity.

- Responsibility for quality check of data rests 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 Municipal Commissioner (*Surat* Municipal Corporation), comprising of CDO, and Additional/Assistant Commissioner may be formed for data quality assessment.

e) SOP for Data Publishing

National Data Sharing and Access Policy (NDSAP) defines standards for publishing datasets and feeds on the open data portal. CDOs must ensure adherence towards defined standards and classification. This SOP describes the steps required for publishing data to the Smart Cities Open Data Portal:

1. **Understand the requirement:** Follow proper procedures to collect the relevant data to be uploaded on the open data portal. Understand the publishing options and the available datasets.
2. **Process the data:** Ensure data is in an appropriate format to be published on the open data portal. It does not contain any personalized information, is open, authenticated and free from defects.
3. **Prepare to deploy/publish data:** Follow procedures specific to the publishing option and work with the appropriate team (when necessary) to publish your data. NDSAP recommends that datasets should be published in an open format and should be machine readable. Data format can be chosen from the list recommended by NDSAP, highlighted in the Appendix.
4. **Publish metadata:** Follow established metadata procedures as per NDSAP and any other guidelines laid down by the city to publish metadata on the Portal and create linkages between data and metadata.
5. **Obtain approvals and finalize deployment:** Obtain the appropriate management approvals for your data based on your selected data publishing option, either from CDO or Municipal Commissioner.

Chief Executive Officer
Surat Smart City Development Ltd.

f) SOP for Engaging Stakeholders

The concept recognizes the value of enhancing engagement among all four stakeholders of the quadruple helix model—Government (Surat Municipal Corporation Departments, Surat Smart City Development Limited and Other Departments of Governments), citizens, academia, and industry, along with improvements in the internal workflow and decision-making processes of city governments.

CDOs along with the team of Data Champions/Coordinators shall assess and document the data requirements of various stakeholders in the city ecosystem, along with frequency of consumption and level of granularity.

Key activities may include:

- Identifying stakeholders from various age groups and ethnicities and engaging them in city initiatives
- Organizing workshops, hackathons/events to promote brainstorming over required 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 data alliances

g) SOP for Data Collection, Processing and Analysis for On-Field Surveys

Field survey is defined as collection and gathering of information at the local level by conducting primary surveys. On-field surveys may be required in situation where data from digital sensors or existing datasets are inadequate. Surveys may be administered to fulfil a certain gap, with a well-defined problem for investigation. Data collected from on-field surveys can help *Surat* Municipal Corporation/*Surat Smart City Development Limited* get a snapshot of how things are at a specific time. The survey research may be descriptive, analytical or evaluative. Field surveys are cost-intensive hence they may be conducted after a thorough mapping of their objectives and outcomes.

Survey Data Processing Steps:

1. **Survey designing:** Survey is a research strategy and not a method. CDOs/concerned team should choose the most appropriate method based on purpose. CDO shall help the concerned team devise suitable surveys/questionnaires by clarifying the objective, determining sample and deciding upon the survey mode to finally create the questionnaire. A research method should not advise the questions, but other way around. Posing difficult to answer questions, in such case a simple rule or scale can be used to help respondents.
2. **Data collection:** Survey can employ a range of methods including questionnaires, interviews or even focus group discussions. Data entry in either format may happen manually or electronically. Going forward, all the e-governance IT applications/systems shall be designed in such a way that manual processes get replaced by automated process without much intervention of humans. As most of process would be automated and handled by an e-mode, data will be available for further analysis.

Chief Executive Officer
Surat Smart City Development Ltd.

3. **Data processing:** Before any analysis is possible, ensuring accuracy and quality of data is paramount. Survey form data is always prone to errors, omissions and other inconsistencies. This data inconsistency and incompleteness, if not edited and corrected on time, can complicate the analysis and may even result in wrong analysis. Data processing shall comprise of various steps necessary for preparing the data for analysis, including editing, data classification, removing redundancies, and preparation of tables. This is an important step when the survey instrument collects qualitative data, which needs to be then represented in a format for analysis.
4. **Data analysis:** Data analysis covers the final step of characterizing and interpreting research findings. In situations where the digital tools are employed for the survey, and the data can be processed easily. Data analysis will involve computation of certain indices or measures along with searching for patterns of relationship that exist among the data groups. The task of analysing quantitative data may be accomplished through statistics. Descriptive statistics is to be used for organising raw data obtained in the process of research, such as tabulation and classification of data. Inferential statistics, also known as sampling statistics, will be used for making inferences or conclusions from the data collected from a sample and drawing generalisations on the entire population.

h) SOP for Data Analysis

Data analysis or analytics is an encompassing and multi-dimensional field that uses mathematics, statistics, predictive modelling and machine-learning techniques to find meaningful patterns and knowledge in recorded data. This will help in using intelligent techniques to uncover actionable insights from the relevant data.

Surat Municipal Corporation/Surat Smart City Development Limited to establish analytical capabilities within the Data Team to accomplish data analysis on the cleaned and processed data. Various examples not narrowing down to the same can be sited as follows:

a) Slice and dice to drill down the data till lowest entity
b) Trend analysis and pattern identification on time series (days, weeks, months, quarter or seasonal, etc.),
c) Trend analysis and pattern identification using various dimensions: Cost, Budget, domain specific parameters etc.
d) Comparison between various parameters in different geographies etc.
e) Visualization to view the trends and patterns for decision making. Converting the data into a more comprehensible and user-friendly format.

Basis the data collected, the Data Team of Surat Municipal Corporation/Surat Smart City Development Limited under the leadership of CDO shall perform various analytical tasks as under:

- Descriptive Analytics: It helps in answering "What is happening?"
- Diagnostic Analytics: It helps in answering "Why did it happen?"
- Predictive Analytics: It helps in answering "What is likely to happen?"
- Prescriptive Analytics: It helps in answering "What should I do about it?"

Chief Executive Officer
Surat Smart City Development Ltd.

8. Annexure

1. OPEN GOVERNMENT DATA (OGD) PLATFORM POLICY

Government of India has listed down policy for external users to access the data hosted on the Open Government Data Platform. This policy can be access at https://data.gov.in/policies. The same needs to be implemented and followed by Surat while implementing the policy document. Further, the Government Open Data Use Licence – India has been approved. This can be accessed at https://data.gov.in/sites/default/files/government_open_Data_Use_licence_india.pdf

2. STANDARDIZATION OF DATA ACCESS AS PER NATIONAL DATA SHARING AND ACCESS POLICY (NDSAP)

NDSAP defines standards for publishing datasets and feeds. These standards need to be adhered to by Surat while sharing its datasets.

Open Source Driven: Datasets are considered to be open by default unless classified as internal, sensitive, protected or restricted.

Metadata: Datasets and apps must be published along with proper metadata. Besides facilitating easy access to datasets, using a common data taxonomy/structure shall be extremely useful in the future for federation/integration of data catalogues. Key Metadata elements for catalogues/resources include:

Catalogues

- **Title (Required):** A unique name for the catalogue (group of resources) viz. Current Population Survey <Year>, Consumer Price Index <Year>, Variety-wise Daily Market Prices Data, City-wise Construction of Deep Tube wells over the years, etc.
- **Description (Required):** Provide a detailed description of the catalogue e.g., an abstract determining the nature and purpose of the catalogue.
- **Keywords (Required):** It is a list of terms, separated by commas, describing and indicating at the content of the catalogue. Example: rainfall, weather, monthly statistics.
- **Group Name (optional):** This is an optional field to provide a Group Name to multiple catalogues 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 to your catalogue.
- **Asset Jurisdiction (Required):** This is a required field to identify the exact location or area to which the catalogue 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 of whether it is 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".

Chief Executive Officer
Surat Smart City Development Ltd.

- **Reference URIs:** 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.
- **Access type:** It mentions the type of access viz. Open, Priced, Registered Access or Restricted Access (G2G).
- **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.

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.).

IF RESOURCE CATEGORY IS APP

- **App type (Required):** It mentions the type of App being contributed viz. Web App, Web Service, Mobile App, Web Map Service, RSS, APIs etc.
- **Datasets Used:** Datasets used for making this app.
- **Language:** Language used for app.

Data Formats: NDSAP recommends that datasets should be published in an open format, which can be accessed without the need for a software licence and should be machine readable. The data could be published in any of the following formats:

- CSV (Comma separated values)
- XLS (Spreadsheet - Excel)
- ODS (Open Document Formats for Spreadsheets)
- XML (Extensive Mark-up Language)
- RDF (Resources Description Framework)
- KML (Keyhole Mark-up Language used for Maps)
- GML (Geography Mark-up Language)
- RSS/ATOM (Fast changing data e.g. hourly/daily)

NDSAP can be accessed at https://data.gov.in/sites/default/files/nDsaP.pdf. Guidelines for implementation of NDSAP are available at https://data.gov.in/sites/default/files/nDsaP%20implementation%20guidelines%202.4.pdf

Chief Executive Officer
Surat Smart City Development Ltd.