Silvassa Smart City:

City Data Policy

Table of Contents:

1. Silvassa Data Policy and Scope

This Policy will apply to all data and information created, generated, collected and archived by SMC / SSCL.

This policy applies to any person/user, organization, administrators, contractors / etc. who intends to access information through open data portal of SMC/SSCL. Specifically, the Data Pol applies to the following information assets of SMC/SSCL

Data /Information that collected, captured, aggregated, processed and shared by SMC
Citizens Data / Information Personnel Data/ Information relating to employees of SMC/ SSCL

Purpose

A data policy is essential to understand the contours of data sharing, privacy, security and ownership in the context of the city. Certain types of data (e.g. an individual's tax payments) are clearly private and should not be shared. On the other band, certain types of data (e.g. air quality sensor readings) are for unrestricted public consumption. Vast amounts of data are in the "grey zone" where clear policies are required that balance privacy, legal and public benefit considerations. Data policy is thus needed to define the contours of collaboration between various governmental/ non-governmental entities on sharing and access of data. The data policy helps to lay out the roadmap of the city in terms of milestones in the adoption of open data, data exchange platforms. Data policy will help answer critical questions regarding data ownership and safety. Another purpose of data policy is to encourage Silvassa city to adopt data driven decision making

2. Data Lifecycle Management

This section, explains the proposed open data life cycle model based. All the processes in the life cycle need to be covered, thus stakeholders can follow the standard process. The six stakeholders involved in this life cycle are top management, information management legal advisor, community manager, data owner, and potential users such as businesses and enterprises, citizens, civil society, developers, and researchers.

30/9/2019

2.1 Data Standards

Data standards are the rules that help keep the publishing and organization of open data orderly and efficient. Open Data sets should adhere to standards in order for it to be comparable, allow analysis, derive insights and interoperable to make it more inclusive. Some data standards should be followed:

- Principles and policy standards
- The Open Definition
- File formats for data
- Domain specific policy standards
- Generic technical standards
- Domain specific technical standards

2.2 Data Categorization

Data will be categorized into two broad categories:

Personal Data: Personal data means data consisting of information which is related to a living 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 City on Data Platform under any data sets. Data sets 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. In other word, anonymization means excluding any personal identifiers from data sets.

2.3 Data Classification

Different types of datasets generated both in geospatial and non-spatial form by different departments/organizations are to be classified as shareable data and non-shareable data. The types of data produced by a statistical system consist of derived statistics like national accounts

30/9/2020

statistics, indicators like price index, data bases from census and surveys. The geospatial data however, consists primarily of satellite data, maps, etc. In such a system, it becomes important to maintain standards in respect of metadata, data layout and data access policy. SMC/SSCL will prepare the negative list of data which will be periodically reviewed by the oversight committee.

Data will be broadly classified into following categories

Class	Definition
Public	Data available for public consumption and use.
Internal Use	Information which could only be disclosed to SMC/SSCL employees for managing operations or delivery of public services on day to day basis.
Sensitive	Sensitive data as defined in various Acts and rules of the Government of India.
Protected	Data which needs to be protected for e.g. Identity of citizens and disclosure /notification needs to be issued by SMC/SSCL in case of any breach or loss of data.
Restricted	Data which could lead to threat to life or loss of public assets or critical infrastructure.

Basic Classification of Data Sets -

- Open by Default: Data sets are considered to be open by default unless classified s internal, sensitive, protected or restricted.
- Meta Data: Data sets and feeds must be published with proper metadata i A. Information about the datasets being published using common data taxonomy/structurere is needed as it helps in providing easy access through Data Platform.
- Data Catalogue: As per NDSAP metadata elements for data sets or feeds is defineded as follows:

- **Title (Required):** A unique name for the catalogue (group of resources) z. Current Population Survey, Consumer Price Index, Variety-wise Daily Mart t Prices Data, State-wise Construction of Deep Tube wells over the years, etc.
- **Description (Required):** Provide a detailed description of the catalogue e.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:** 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 (i.e.) 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.

- **Open data:** Data Sets and feeds should be published in formats as per guidelines of NDSAP i.e. Open format. Data should be provided in freely available formats which can be accessed without the need for a software license.
- **Machine Readable:** Data Sets and Feeds should be machine readable.

Formats: following data formats to be published:

- CSV (Comma separated values)
- XLS (Spread sheet - Excel)
- ODS (Open Document Formats for Spreadsheets)
- XML (Extensive Markup Language)
- RDF (Resources Description Framework)
- KML (Keyhole Markup Language used for Maps)

30/9/2019

- GML (Geography Markup Language)
- RSS/ATOM (Fast changing data e.g. hourly/daily)

- **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 through social media platforms like Facebook, Twitter.
- **Ownership:** All data sets/ feeds remain property of SMC/SSCL. The. CDO will endorse government open data license to ensure that published data is not misused misinterpreted by its users.

2.4 Data Archival and Retention

CDO set up process for archiving process. Every data set/feeds catalogue contains archiving information. for specific file type (geo-spatial files), recent copy will be made available to users through Data Platform.

E-Files/records will be digitized by the following categories:

a) Category-I (e-Files/records to preserved permanently which are of historic 1 importance) - For 5 years, it will be kept in the Department's sever and thereafter transferred to Data cloud storage.

b) Category -II (e-Files/records of secondary importance and have a reference value i or a limited period} - 5 years on the Department's server_ In exceptional cases, if he records is required to be retained beyond 5 years It will be upgraded to category-I.

Data will be stored in the main database for 6 Months in a live Mate so that when a report needs to be generated, the data will be extracted from main database_ I), ♦ to older than 6 months will be archived. if report duration extends beyond 6 months, the. data will be retrieved from archival to generate the report.

2.5 Data Security and Privacy

Data Security and Privacy will be managed by SMC/ SSCL under the purview of the Data Policy.

30/9/2022

a) **Physical Security**

The premises will be physically secured, access control devices Should available for accessing the premises, entry and exit should be monitored.

b) **Network Security**

Appropriate firewalls, IPS, SSL devices etc. should be used to ensure Network security.

The solution should support encryption mechanism for transferring data across network and between client and server.

c) **System Security**

- Adequate access control procedures should be followed to secure the entire IT system, physically and logically.
- The access controls procedures should cover all stages in the life-cycle of user access, from the initial registration of new users to the final de-registration of users who no longer require access to information systems and services.
- The system should have 2 factor authentication mechanism either through One Time Password **(OTP)** or soft token - based technologies for access control and user authentication.

d) **Application Security**

- The solution should have appropriate authentication mechanisms.
- Application user authentication & authentication related transactions should be encrypted.
- Operating system should be hardened on which the application is installed.
- A web application firewall shall be deployed to secure the web layer.

e) **Audit Trails & Logs**

- Event logging should create an accurate record of user activity such as which users accessed which system, and for how long.
- The solution should log all types of events especially those related to security.

f) **Data Protection**

- The solution should support SSL encryption mechanism for transferring data across network.
- The data transferred across network should be encrypted 'using (PKI)Public Key Infrastructure.
- 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 order to provide more effective protection for an electronic document in order r o maintain Confidentiality, Authorization, Accountability, Integrity, Authenticity and Non-repudiation.

g) **Session Management**

- The system should limit to only one session per user or process **ID.**
- The system should put a limit on the maximum time length of an idle session, which should ensure that automatic session termination takes place after expiry of the specific time length.
- Mandatory password change after predefined time period.

h) **Data Warehouse Security**

- Users must not have access to the data warehouse prompt of the application Access to the data warehouse prompt must be restricted only to the database administrator.
- "Super user" rights for the data warehouse must only be given to t administrator and activities of these accounts must be properly logged.

i) **Application Deployment**

- All unused ports should be blocked at server machines.
- The application server should be segregated from Internet zone through firewall or other filtering

30/9/2020

J) **Information Security Governance**

- The employees working on the project should be made aware of his or her responsibilities with respect to Information Privacy and Information Security.
- Employees working on the project shall undergo security awareness trains fig during training.

k) **Compliance to Security Standards**

- Software/Hardware system should be in compliance with ISO/r,C 27001:2015.

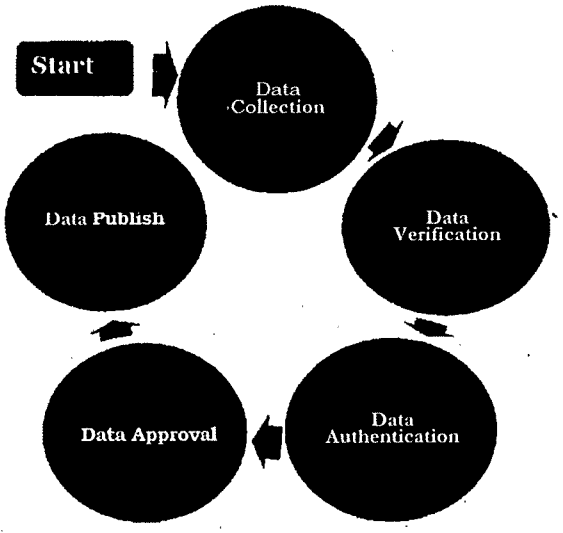
Security Information and Event Management System (SIEM)

- SIEM should be available for Real-time analysis of security alerts generated by applications and infrastructure.

Database Activity Monitoring (DAM)

DAM should be available to monitor all database

Data Flow Chart



2.6 Roles and Responsibilities

City Data Officer (CDO)

The CDO will be the officer responsible for implementation of this data strategy at the city level.

The key responsibilities of CDO is as follows:

1. The CDO will form City Data Policy (CDP) which will be reviewed every month to keep it contextual to the need of the times. The CDA would act as advisory body for the review of CDP from time to time.

2. Coordination with MDO (Mission Data Officer) to align with mission data strategy and priorities

30/9/2020

with respect to Open government initiatives and policies.

3. Organize regular meetings of CDA (City Data Alliance)

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

5. Publish Data Catalogues and Data Sets/Feeds on OGD portal: CDO will publish data Catalogues and Data Sets/Feeds on OGD Portal and will ensure that such data sets are updated at regular time intervals as needed and create mechanisms, for continuous feedback from citizens and stakeholders on type of data sets to be published .

Data Champions:

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

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

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

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

DCs needs to act as trainers and lead the team of data coordinators at the department le\ el. It is critical for appropriate senior functionaries to be designated as DCs in ew h department by the respective city heads. The DC will the nodal point for implementation, of the CDP within the department and will function to supervise the team of dc, is coordinators on day to day basis. DC will be first touch point of CDO in different city organizations in the city and must undertake continuous capacity building

30/9/2019

programs and their DCOs and other staff.

Data Coordinators:

a. Data Coordinators will assist Data Champions at the department/government agency level as reporting staff

b. Data Coordinators will also aggregate the data demand from various channels.

c. Data Coordinators will also be responsible for sensitizing the department employees over the importance of data quality etc

d. Data Coordinators shall perform collection, interpretation and recording of data in accordance with City Data Policy standards and CDO guidelines

e. Data Coordinators shall perform data validation and storage of various project documents.

f. Data Coordinators shall review required data and documents & make necessary revisions to the same

g. Data Coordinators shall Sort and organize the data; both hard copy and electronic versions.

h. Transmit data report to SMC/SSCL or CDO via Internet

i. Update SMC/SSCL website or Silvassa Open Data Portal with latest data records

j. Data Coordinators shall maintain the completed hard copy and electronic files of project records.

k. Data Coordinators shall assist department staff in data entry when required. 1. Data Coordinators shall provide data management updates in all internal and external meetings as required.

l. Data Coordinators shall analyse data for quality improvement purposes.

m. Data Coordinators shall Prepare data for reporting, meeting and presentations for their department and SMC/ SSCL at large.

n. Data Coordinators will ensure data management procedures comply with City Data policy

o. Provide statistical analysis and longitudinal analysis of data

30/9/2020

Prepare and submit data required for audits.

City Data Alliance (CDA):

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

The key roles for setting up City Data Alliance (CDA) are as follows:

To act as an advisory group to the city leadership on the City Data Policy

1) To promote data driven governance and policy formulation.
2) To design and implement solutions and analysis using city data.
3) To support industry to design solutions using emerging technologies like AI, ML and Blockchain.
4) To assess and design use cases critical to the citizens of the city.
5) To generate awareness in various stakeholders towards open government initiatives. To facilitate data for co-creation and collaboration over civic issues
6) To provide critical feedback to the city over the quality and relevance of data provided by City.
7) To deliver 4 Research paper annually using City Data on Civic Problems in the City) To design and develop two prototype/ solutions annually on Civic Problems the City '() To organize a data, challenge every half yearly on complex civic problems
8) To organize a Hackathon annually and support shortlisted solutions at city level

30/9/2020

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

10) To publish the progress, report every month

11) Prioritize the Data Sets/Feeds for publishing on Data Platform:

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

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

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

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

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

2.7 Standard Operating Procedures:

SoP for Manual Data Collection

1. The request received can be directed to the Data Coordinator of the concerned department.
2. The Data coordinator of the concerned departments checks the requested data. If data is available with the department, DC shall instruct the concerned personnel to gather the data in the requested format.

If fresh data is needed to be captured / acquired, the Data Coordinator (DC) in consultation with City Data Officer shall take appropriate actions.

30/9/2020

SoP for Electronic Data Collection

1. The data should be collected in consent with the end-user who may be a citizen or SMC/SSCL employee
2. Data which is not going to be used for any kind of analysis or will not be used for al y communication purpose should not be collected at all.
3. Data Security measures mentioned in "Data security' section of the Policy, shall be followed to maintain confidentiality and security of data
4. For Data Collection and Integration with IOT systems / Sensors, the following Data Architecture shall be followed by (SMC) / (SSCL)

Data Collection

a) Field Data: SMC/SSCL should deploy various IoT sensors, actuators, devices, cameras and solutions to capture the data from the field directly.

b) Operations Data: SMC/SSCL should deployed various system under various government departments to manage city core operations like Water Supply, Surveillance, Traffic Electricity, Street Lights, Water and Sewerage Treatment, Health, Education, Fire Department, Disaster Response and License permits etc. The data Generated by these systems in different formats should provide critical information and intelligence to officers to manage critical services and Infrastructure of the city. There should be arrangements to collect, store, analyze and generate meaningful insights from this data.

c) Third Party Platform and Mobile Apps: Arrangements to be made to avail services from private businesses like radio taxi, food delivery, hospitals and labs etc. which could provide various insights and patterns which could be instrumental in policy formulation and city planning.

d) Internet: Various platforms engages citizens and communities which captures general sentiments which could provide insights mood or opinion of citizens and communities towards specific issue. There should be arrangements to collect, store, analyze and generate meaningful insights from this data.

Data Integration

Data from different systems using different technologies comes in different size, shape format. In order to derive meaningful information from structured and unstructured data it is required to make data compatible for consumption.

Data Analysis

Analytics is an encompassing and multidimensional field that uses mathematics, statistics, predictive modelling and machine-learning techniques to find meaning) In patterns and knowledge in recorded data. Applying intelligent techniques to uncover insight from the relevant data. SMC/SSCL to set up Analytics division to accomplish Data Analysis insights from the Data sourced and integrated

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 patter 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.

5. **4.** Data Intelligence

It is final stage of the journey from being informed to actionable insights and finally to actions using following techniques:

The Analytics team of SMC/SSCL under the leadership of CDO shall perform (not limited to) following basic analytical tasks

- **Descriptive Analytics:** It helps in answering "What is happening?" For Example: Using past financial performance to predict a customer's likely financial performance. Descriptive analytics can be useful in the sales cycle, for example, to categorize customers by their likely product preferences and sales cycle.

- **Diagnostic Analytics:** It helps in answering "Why did it happen?" For Example: For a social media marketing campaign, you can use descriptive analytics to assess the number of posts, mentions, followers, fans, page views, reviews, pins, etc. There can be thousands of online mentions that can be distilled into a single view to see what worked in your past campaigns and what didn't.

- **Predictive Analytics:** It helps in answering "What is likely to happen?" For Example: Some companies are using predictive analytics for sales lead scoring. Some companies have gone one step further use predictive analytics for the entire sales process, analyzing lead source, number of communications, types of communications, social media, documents, CRM data, etc. Properly tuned predictive analytics can be used to support sales, marketing, or for other types of complex forecasts.

- **Prescriptive Analytics:** It helps in answering "What should I do about it?" For Example: In the health care industry, you can better manage the patient population by using prescriptive analytics to measure the number of patients who are clinically obese, then add filters for factors like diabetes and LDL cholesterol levels to determine where to focus treatment. The same prescriptive model can be applied to almost any industry target group or problem.

20/9/2019