Smart City

SAHART

City Data Policy

SAHARANPUR
Municipal Corporation

Prepared By:

Saharanpur Smart City Limited (SSCL)

Table of Contents

Abbreviations

SMC	Saharanpur Municipal Corporation
SSCL	Saharanpur Smart City Limited
SPV	Special Purpose Vehicles
ICCC	Integrated City Command and Control Centre
ITMS	Intelligent Traffic Management System
OGD	Open Government Data
CDP	City Data Policy
CDO	City Data Officer
SCDA	Smart City Data Alliance
DC's	Data Champions
DCO's	Data Coordinators
CSV	Comma - Separated Values
XLS	Excel Spreadsheet
ODS	Open Document Format for Spread sheets
G2G	Government-to-Government
G2C	Government-to-Citizen
G2B	Government-to-Business
API	Application Programming Interface
OGD	Open Government Data
SSL	Secure Sockets Layer
SCADA	Supervisory Control and Data Acquisition
NDSAP	National Data Sharing and Accessibility Policy
NIC	National Informatics Centre
GOI	Government of India

Introduction

The importance of data democratization and open data in a world where data is touted as the new policy is unparalleled. Open Government Data refers to a set of policies which facilitates availability of government datasets (administrative, transactional, and service delivery) in the public sphere to enhance transparency and accountability in the ecosystem. It is expected to trigger a higher quotient Client of trust towards governments, increases collaboration and engagement with citizens, and private entities; and 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 about the steps and processes that need to be put in place to enhance impact.

SMC / SSCL strives to make the data it generates and collects openly available to the public for the purposes of increasing the quality of life for our residents; increasing transparency, accountability and comparability; promoting economic development and research; and improving general performance management. This Open Data Policy defines the principles governing City data and describes the expectations for department participation and governance of the City Data Program for Saharanpur.

The proposed Open Data Guidelines are intended as a resource for city administrators such as Municipal Commissioner and Smart City CEO, other officials such as City Data Officer, Heads of various government departments, data coordinators and data champions; and external agencies, civic, private interested in partnering with the data initiatives for the City.

1. Definitions:

- **Data-** Data means a representation of information, numerical compilations and observation, documents, facts, maps, images, charts, tables and figures, concepts in digital and/or analog form.
- **Open Data** — Specific Datasets that are made available for the public of the city.
- **Data Archive** —Data archiving is the process of moving data that is no longer actively used to separate storage device for long-term retention. Archive data consists of older data that remains important to the organization or must be retained for future reference or regulatory compliance reasons. Data archives are indexed and have search capabilities, so files can be located and retrieved.
- **Data Generation** -Data generation refers to the theory and methods used to create data from a sampled data source in a qualitative manner. Data sources include human participants, documents, organizations, electronic media, etc.
- **Data set** -A data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in a question .
- **Non-sharable data-** Data which are confidential and non-sharable as declared by the departments / organizations.
- **Sensitive data** - Sensitive data as defined in various Acts and rules of the Government of India.
- **Sharable data-** Data which are easily access to public and not covered under any restriction.
- **Restricted Data** — Data which are accessible only through a prescribed process of registration and authorization by respective departments / organizations.

मुख्य कार्यकारी अधिकारी (सी०ई०ओ०)
सहारनपुर स्मार्ट सिटी लिमिटेड

2.Need of Data Policy for Saharanpur

City Government departments generate a large amount of data. This data is generated as part of daily activities of the municipal corporations and city governance department. The need to facilitate sharing and utilization of this large amount of data primarily points to the need to a skeleton structure defining rules and regulations. Today most of the government information is stored in databases, the notion that information created in the public interest belongs to the public fuels the development of modern urban civics. 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 sharable data available with data owners. Such regimes could lead to duplication of efforts and loss of efficiency of planning of activities focused on development.Hence, City Data Policy of Saharanpur Municipal Corporation aims to provide an enabling provision an platform for providing proactive and open access to the data generated through public funds& public revenue available with various departments of SMC/SSCL, other government departments etc.

3. Saharanpur 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 policy applies to the following information assets of SMC/SSCL:

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

3.1 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 hand, 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 Saharanpur city to adopt data driven decision making.

3.2 Data Policy Goals:-

Following are the Open Data Policy Goals for the City of Saharanpur.

a) Increased Transparency and Accountability thus fostering greater trust on government.
b) Increased public participation in government data analysis and deliver solutions or ideas for betterment of city governance.
c) Improved resource or asset visibility, social audit and open government
d) Better decision making thereby leading to more efficient and cost-effective solutions.
e) Deepen open innovation and co- creation.
f) Foster data driven decisions by diverse players in urban economic ecosystem

g) Foster advanced research in academic and research institution
h) Helps cities develop new business models
i) Empowers communities through sharing of data
j) Promotes development of emerging technologies like AI, ML and Block chain
k) Enhanced Government to Government(G2G), Government to Business (G2B) and Government to Academia (G2A) collaboration

4 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 manager, legal advisor, community manager, data owner, and potential users such as businesses and enterprises, citizens, civil society, developers, and researchers.

4.1 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.

4.2 Data Categorization

Data will be categorized into two broad categories:

a) 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 anonym zed before publishing.

b) 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.

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 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 -

A) Open by Default: Data sets are considered to be open by default unless classified as internal, sensitive, protected or restricted.

B) Meta Data: Data sets and feeds must be published with proper metadata. Information about the datasets being published using common data taxonomy/structure is needed as it helps in providing easy access through Data Platform.

C) Data Catalogue: As per NDSAP metadata elements for data sets or feeds is defined as follows:

- **Title (Required): A** unique name for the catalogue (group of resources) .

  Current Population Survey, Consumer Price Index, Variety-wise Daily Market Prices Data, State-wise Construction of Deep Tube wells over the years, etc.
- **Description (Required):** Provide a detailed description of the catalogue, 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.

D) 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.

E) Machine Readable: Data Sets and Feeds should be machine readable.

F) 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)
- GML (Geography Markup Language)

- RSS/ATOM (Fast changing data e.g. hourly/daily)

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

H) 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 Face book, Twitter.

I) 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.

4.4 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 i.e. 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 .if 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 for a limited period} - 3 years on the Department's server In exceptional cases, if lie record 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 whenever a report needs to be generated, the data will be extracted from main databases Data older than 6 months will be archived. if report duration extends beyond 6 months, th e data will be retrieved from archival to generate the report.

4.5 Data Security and Privacy

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

a) **Physical Security**

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

b) **Network Security**

- Appropriate firewa1ls, 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 mechanisms 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 in order to provide more effective protection for an electronic document in order to 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 the 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.

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 training during training.

k) Compliance to Security Standards

Software/Hardware system should be in compliance with ISO/IEC 27001:2O15.

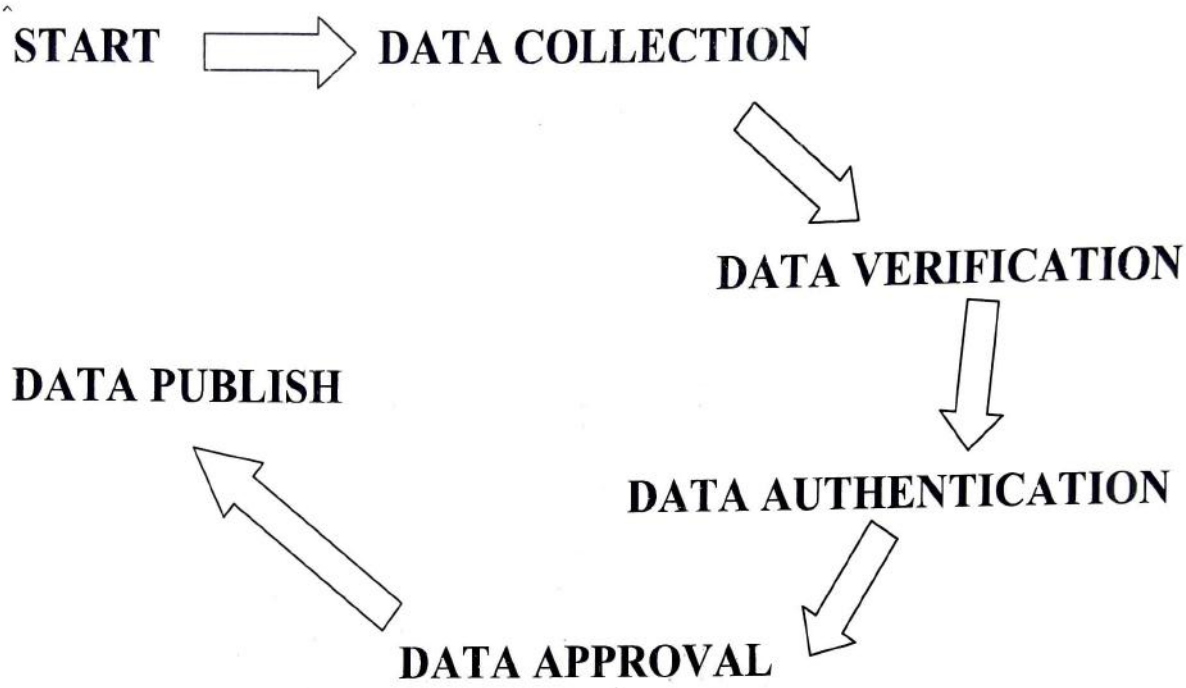
l) Security Information and Event Management System (SIEM)

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

m) Database Activity Monitoring (DAM)

DAM should be available to monitor all databases.

4.6 Data Flow Chart:



5.Data Provisioning and Identifying Datasets:-

Focus on identifying high value datasets that would have direct impact on citizen welfare will (in terms of city infrastructure and services) and the overall wellbeing of citizens. Based on the utility of these data sets, SMC/SSCL could substantiate cases for garnering additional technical support and funding from state and central government. This set can serve two purposes - as a primary set of datasets to begin the initiative (basic datasets), or as a reference set of datasets which SMC/SSCL should consider to keep adding datasets (Intermediate or Advanced datasets) to the Open Data portal.

- **Demand of Public Data Sets -** Like any other product or service offering, demand of the data set should be one of the key drivers while deciding about making it public. Key focus groups influencing this demand should be identified and engaged with during at an early stage for their inputs on a particular datasets and the priority with which it is required. In some cases SMC/SSCL could present a list of datasets that can ask the focus group on data that would meet public need.
- **Social Impact** — Datasets which potentially could create a positive social impact, irrespective of the economic value generated, should be included. Such positive impacts can include generation of employment, boosting equality and equal opportunities among citizens, address social issues like improving female student enrolment in primary and secondary schools, and empowerment of the socially backward or oppressed classes etc.
- **Economic Value Generation** — Publishing certain datasets can boost innovation, encourage entrepreneurship eventually resulting in generation of economic value i.e. Greater economic value could have a far-reaching impact — including generation of employment, boosting per capita income, and bringing overall prosperity for the city and the nation. SMC/SSCL should identifying and prioritizes Open list datasets (datasets which are not part of negative list) into 'high value' datasets and 'non-high' values datasets on basis of dataset's potential to generate economic value.
- **Legal and Compliance requirement** — There can be legal and compliance requirements for mandatory disclosure of a certain dataset to the public. SMC/SSCL should proactively identify datasets, publishing which would help reduce volume of RTI queries significantly, and could result in lowered burden on SMC/SSCL (and other government bodies in the city).
- **Minimal Resistance** — Opening up data sets can have huge implications politically as it could potentially unearth issues and problems that could take a political any once the data is made public. SMC/SSCL needs to be conscious on how data represented and made accessible, so it can

be utilized by various parties.

6. Stakeholders and Collaboration

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

a) **Government Agencies:** Government Agencies operating with dedicated administrative structure 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 works with city administration in different domain for e.g. World Bank, ADB, DFID, AFB etc.

c) **Industry:** Key flagship manufacturing/service Industry promoters/players in the 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:** Representatives from start-ups in the City/State.

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

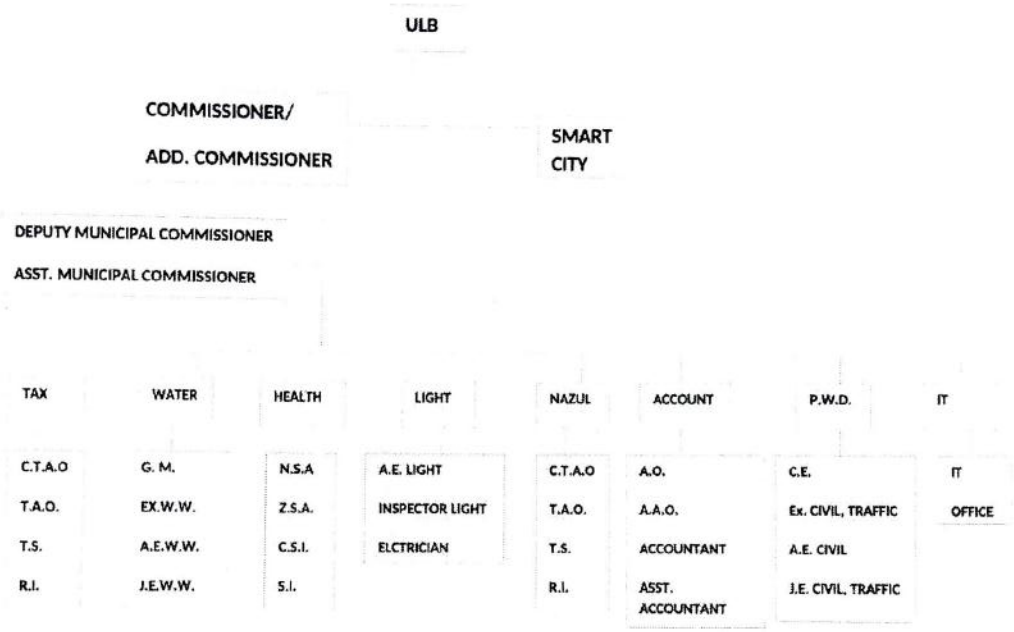
h) **Citizens and Communities:** Representatives from Communities and citizen interest groups to further the interest of citizens/communities towards data driven policy governance and service delivery.

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

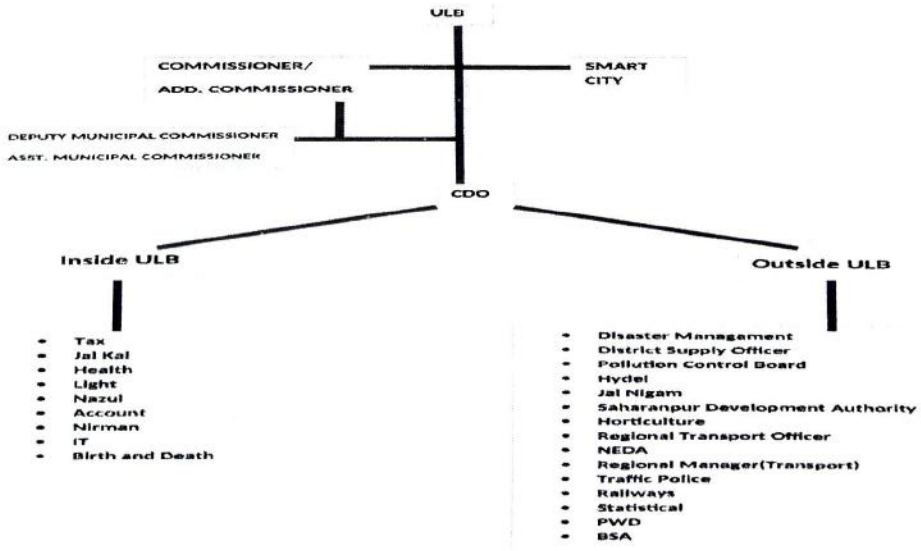
j) **Professional Representatives:** Representatives from various professional bodies like Doctors, A and Engineers etc.

7. Data Management and Team Structure

City Data Committee should be formed to engage participation and secure buy-in from both internal and external stakeholders on key decisions. Further, sufficiently empowering this committee can help navigate through complicated hurdles (e. bureaucratic, political etc.) and to prompt decisions and actions pertaining collection, segregation and release of open data.



EXISTING STRUCTURE



FUTURISTC STRUCTURE

8. Roles and Responsibilities:

8.1 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 are 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 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 .

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

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

8.3 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. Data Coordinators shall Sort and organize the data; both hard copy and electronic versions.
g. Transmit data report SMC/SSCL or CDO via Internet
h. Update SMC/SSCL website or Saharanpur Open Data Portal with latest data records
i. Data Coordinators shall maintain the completed hard copy and electronic file of project Records.
j. Data Coordinators shall assist department staff in data entry when required.
k. Data Coordinators shall provide data management updates in all internal and external meetings as required.
l. Data Coordinators shall analyze data for quality improvement purposes.
m. Data Coordinators shall prepare data for reporting, meeting and presentations for their department and SMC / SSCL atlarge.
n. Data Coordinators will ensure data management procedures compliance with City data Policy.
o. Provide statistical analysis and longitudinal analysis of data.
p. Prepare and submit data required for audits.

8.4 City Data Alliance (CDA):

CDA for the city is envisaged to be network of government departments, agencies, private sector 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 defined 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
- To promote data driven governance and policy formulation.
- To design and implement solutions and analysis using city data.
- To support industry to design solutions using emerging technologies like AI, ML and Block chain.
- To assess and design use cases critical to the citizens of the city.•
- To generate awareness in various stakeholders towards open government initiatives.
- To facilitate data for co-creation and collaboration over civic issues
- To provide critical feedback to the city over the quality and relevance of data provided by city
- 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
- To organize a Hackathon annually and support shortlisted solutions at city level
- To set up scholarship for postgraduate and graduate interns to work with Office of CDO. To publish the progress, report every month.

- Prioritize the Data Sets/Feeds for publishing on Data Platform
- To sensitize ecosystem partners to share the data for leveraging data for solving civic challenges
- To support, engage and encourage network/groups/members of data enthusiasts in the city.
- To improve city capacity over data driven governance and policy formulation.
- To support CDO by extending resources (like interns, researchers, technology experts), funds (program sponsorship etc.) and technology (solutions etc.)
- To share data available with partners on Data Platform to promote City Data.

9. City Data Policy: - Implementation Plan

Open Source Driven-Developed completely using open source stack, facilitating cost saving in terms of software and licenses and also provisioning community participation in terms of further development of product with modules of data visualization, consumption, APIs to access datasets.

- **Metadata-** Resources shall be published along with standard metadata along with controlled vocabularies on government sectors, jurisdictions, dataset type s, access mode etc. Besides facilitating easy access to datasets, this shall be extremely used al in the future for federation/integration of data catalogs.
- **Social Media Connect**—To support wider reach and dissemination of datasets, anyone can share the information about any dataset published on the platform with his/her social media pages.
- **Citizen Engagement** — The Platform has also a strong component of Citizen Engagement. Citizens can express their views as well as rate the datasets w.r.t three aspects (Quality, Accessibility and Usability) on the scale of 5. They can also embed the Resources (Datasets/Apps) in their blogs or web sites. Facility to contact the Chief Data Officers are also available on the Platform.
- **Community Collaboration** — Citizens with specific interest can build communities and discuss online. OGD Platform facilitates the communities to open up online forums, blocks and discussions around various datasets, apps available on the platform. It also provide a platform to express and discuss the kind of Datasets, APPs & APIs they would like have. It shall also give input to departments as what kind of datasets is more useful an accordingly prioritize the release of the datasets.

10. Standard Operating Procedures: -

10.1 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.
3. If fresh data is needed to be captured / acquired, the Data Coordinator consultation with City Data Officer shall take appropriate actions.

10.2 SoP for Electronic Data Collection

1. The data should be collected in consent with the end-user who may be a citizen for SMC/SSCL employee.
2. Data which is not going to be used for any kind of analysis or will not be used for 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).

The City reference data architecture is explained below:

A. **Data Sources:** This ecosystem comprises of various technology solutions ranging from Sensors, IoT, SCADA, Electronic camera, GIS, payments system etc. which generates loads of structured data every second on different dimensions.SMC/SSCL ecosystem could leverage various unstructured data emerging from different sources and third-party systems like social media, Internet, websites, videos, images etc.

B. **Data Collection and Acquisition:** Data Collection and Acquisition Layer acts as an interface between data sources and Intelligent Platform (Command and Control Platform, Analytics Platform) which comprises of engines and solutions to customize and prepare the data collected through various sources and formats i.e. structured and unstructured for analysis purpose. (Refer Data Integration in point 5).

C. **Enterprise Data Repository:** Processed Structured and Unstructured Data is stored in enterprise Data Warehouse from where it could be used by various users and applications for decision making.

D. **Data Analysis Layer:** Comprises of engines to process the structured and unstructured data on various dimensions for various purpose. Data Analysis Layer engine helps user to derive information, intelligence and knowledge out of processed data stored in Enterprise Data Repository for analysis or decision-making purpose. (Refer Data Analysis in points i.e. Lifecycle of Data.)

E. **Service Delivery Layer:** Service Delivery layer comprises of various applications and systems which could be used to deliver information, intelligence and knowledge to end user.

5. Life-Cycle of Data: Explained

5.1. 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 engage 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.

5.2 Data Integration

Data from different systems using different technologies comes in different size, shape .and format. In order to derive meaningful information from structured and unstructured Data it is required to make data compatible for consumption. Various data types and form it's generated from various systems are as follows:

A. Field Sensors and Devices:

- Hierarchical files (JSON, XML, YAML, etc.)
- Real-time stream
- Objects

- Videos
- Images
- Location Data

B. Operations Systems:

- Relational Data Structures
- Blocks of raw data
- Flat text files
- Documents (xls.pdf, put, etc.)
- Log files
- Financial Data
- Location Data

C. Internet and Social Media Content:

- Blogs
- Video
- Music
- Sentiments
- Images

Under City Data Policy arrangements should be made to build a data warehouse and integrate data with it. Extract, Transform and Load (ETL) is the common methodology used for data integration. It is a three step process which used for data integration to blend data from multiple sources. 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.

5.3- Data Analysis

Analytics is an encompassing and multi dimensional field that uses mathematics, statistics, predictive modeling and machine learning techniques to find meaningful pattern 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 follow

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.

10.3 SoP for Data Processing and cleaning

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 of concern authority.

10.4 SoP for Data for Quality Assessment and Data Sets

1. Under the leadership of Municipal commissioner (SMC), three-member committee will be formed comprised of Additional Commissioner (SMC). Assistant Commissioner(S), Chief Accountant Officer.
2. Chief Data Officer (CDO) and Data Coordinators shall submit their data to the committee for Data quality assessment.

10.5 SoP for Data Publishing as per Open Data Norms

Saharanpur City Data Policy endorses National Data Sharing and Access Policy which defines following standards for publishing data sets and feeds.

1. Only data which has been approved by Assessment Committee and Municipal Commissioner shall be uploaded on Open Data Portal.
2. Data set which is considered to be open by default unless classified as Internal, sensitive, protected or restricted shall be uploaded on the Open Data Portal.
3. Data sets and feeds must be published with proper metadata. Information about the datasets being published using common data taxonomy/structure shall beuploaded, is it helps in providing easy access through Data Platform.
4. A Data Sets and feeds should be published in formats specified under NDSAP open format. Data should be provided in freely available formats which can be accessed without the need for a software license.
5. Data Sets and Feeds should be machine readable.
6. Following data formats can be used for uploading data on Open Data Portal
   - **A.** CSV"(Comma separated values)
   - **B.** XLS (Spread sheet - Excel)
   - **C.** ODS (Open Document Formats for Spreadsheets)
   - **D.** XML (Extensive Markup Language)
   - **E.** RDF (Resources Description Framework)
   - **F.** KML (Keyhole Markup Language used for Maps)
   - **G.** GML (Geography Markup Language)
   - **H.** RSS/ATOM (fast changing data e.g. howdy/daily)

10.6 SoP for Engaging Stakeholders to Assess the Data Needs

Data sharing between all stakeholders is the only way which can catalyze collaborative problem solving. Various ways of collaboration around data sharing to drive open innovation shall comprise of the following

1. API: Stakeholders (Internal / External / citizens / Public or Private limited entities/ Registered NGOs) should strive to automate the extraction of data and illustrate trends easily and quickly with APIs.
2. Data Pooling: Data is provided by various data producers from across the city governance. SMC/SSCL can take this data and organizes it on its website by category, thus helping Site visitors to see the history of data available, compare two set of data, or view data by source.
3. Intelligence Product: Data to be made readily available to researchers and external agencies, who can use existing data to create predictive models and machine learning techniques to guide data-driven services for SMC/SSCL
4. Prizes and Challenges: Public data will be used for the challenge and available through some data platform, Participants will also be encouraged to include data from their own organizations and add this data to the Data Registry.
5. Research Partnerships: SMC/SSCL to establish research partnership with researchers through a secured protocol to avoid any data leakage or misuse.
   Trusted Intermediary: SMC/SSCL can buy data from trusted Intermediary (eg: Google, Face book, LinkedIn, social media, etc) for better decision making and city governance.

10.7 SoP for Data Collection, Processing and analysis for on field survey

Depending on requirement of data, competent agency can be employed to perform field survey. Going forward, all the e-governance IT applications/ Systems shall be designed in such a via 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.

10.8 SoP for Data Monetization

A committee comprising of representatives from various departments of SMC/SSCL, Traffic Police, Uttar Pradesh Environment Conservation Board, Saharanpur Police, city educational institutions and market experts would be formed to be City Data Committee.

In the context of Right to Information act, this committee would decide the data which can be monetized, fix the price, and review it from time to time. The data can be differentially priced for academic research and commercial used.

SMC will create data monetizing platform, it will form revenue sources that would help the data ecosystem to become self-sustainable. It will allow SMC to broker third party data and benefit from this brokerage.

10.9 Processes defined including Provisions for Data Analysis

Apart from presenting the actual data, cleaned datasets might also be used to discover new patterns or analyse existing patterns, trends or behaviors. Data analysis may help multi-disciplinary researchers provide different perspectives or even solutions on civic issues like transport, traffic, solid waste etc. In line with this, the SOP should define the contours of analysing the collated data. It should shed light on how the data could be utilized to provide meaningful analysis that can help with solving urban problems.

Data analysis comprises of tools and methods used to process structured and unstructured data on various dimensions for various purposes. This will help users derive information, intelligence and knowledge out of processed data.
Further emerging technologies such as analytics, artificial intelligence, machine learning etc. may be used for generating insights from the data gathered from the multitude of sources which will help in improved decision making. For example:

• Trend analysis and pattern identification on time series data (days, weeks, months, quarters or seasonal etc.)
• Trend analysis and pattern identification using various dimensions: cost, budget, domain-specific parameters etc.
• Comparison between various parameters in different geographies etc.
• Visualization to view the trends and patterns for decision making, by converting the data into a more comprehensible and user-friendly format.

11. Top Priority Areas for City Data Policy Implementation

- The IT Department of SMC would be serving the nodal functions of coordination and monitoring of policy through close collaboration with all Central Ministries and i tie Department of Information Technology by creating the web portal.
- All sharable data as decided by the City Data committee will be made available on 'ar.-is where-is' basis.
- The IT Department of SMC will make sure the implementation is in compliance with detailed implementation guidelines including the technology and standards for data and metadata to be brought out by Department of Information Technology, Government of India.
- All the data users who are accessing/using the data shall acknowledge the source department in all forms of publications.
- All SMC Departments will upload at least high value datasets decided by the City Data

Committee on the Saharanpur data portal within three months of the notification of the police

- Uploading of all remaining datasets should be completed within one year by all If relevant departments.
- Saharanpur Data Portal itself will have both data and Meta data which could be accessed from the portals of the SMC departments.
- All metadata will follow standards and will minimally contain adequate information proper citation, access, contact information, and discovery. Complete information including methods, structure, semantics, and quality control/assurance will be follow at for most datasets.
- SMC will design and position a suitable budgetary incentive system for data owners for increasing open access to the sharable data under guidance of the Government of Uttar Pradesh.
- An oversight committee will be constituted for facilitating the implementation of the policy and its provisions thereof
- The IT department of SMC will constitute a coordination committee for facilitating implementation of the policy and its provisions thereof.

12. Amendments to Policy

- Technology and database ecosystem are a highly dynamic field. Its needs and challenges keep on evolving with time to keep this data policy relevant and appropriate with time, every year the City Data Committee would review the policy in consultation with subject matter experts and would be empowered to amend the policy.

मुख्य कार्यपालन अधिकारी (सी०ई०ओ०)
सहारनपुर स्मार्ट सिटी लिमिटेड
सहारनपुर

Annexure:1

Resource/Dataset	Catalog/Resource Name for Open Data Portal
Dataset 1	Demographic Profile:<Your City Name>
Dataset 2	Unemployment Rate:<Your City Name>
Dataset 3	Household Profile :<Your City Name>
Dataset 4	Air Quality S02,PM2.5,PM10,NO@,O3: :<Your City Name>
Dataset5	Public Amenities:<Your City Name>
Dataset6	Community Facilities:<Your City Name>
Dataset 7	Citizen Facilitation Center<Your City Name>
Dataset 8	Health Infrastructure Catalog<Your City Name>
Dataset 9	Mortality Rate<Your City Name>
Dataset 10	Water Revenue Data<Your City Name>
Dataset 11	Public Toilets Data<Your City Name>
Dataset 12	D2D Collection Coverage<Your City Name>
Dataset 13	Solid Waste Generated, Collected, Processed Data<Your City Name>
Dataset 14	Solid Waste Ward Wise Segregation<Your City Name>
Dataset 15	Solid Waste Collection Revenue Data<Your City Name>
Dataset 16	Bins And Community Bins Ward Wise Data :< Your City Name>
Dataset17	Waste Collection Vehicle Data<Your City Name>
Dataset 18	Solid Waste Processing Data<Your City Name>
Dataset 19	Solid Waste Disposal Data<Your City Name>
Dataset 20	Waste Management Revenue Expenditure Data :< Your City Name>
Dataset 21	Property Tax Data<Your City Name>
Dataset 22	Education data :< Your City Name>
Dataset 23	Condition of roads<Your City Name>
Dataset 24	Street lights<Your City Name><Your City Name>
Dataset25	Slum Housing and population data<Your City Name>
Dataset26	Housing and Basic Infrastructure Data :< Your City Name>

Dataset 27	Circle Rate :< Your City Name>
Dataset 28	Infrastructure Distribution :< Your City Name>
Dataset 29	Open Spaces :< Your City Name>
Dataset 30	Natural Landscape<Your City Name>
Dataset 31	Cultural Heritage<Your City Name>
Dataset 32	Children Facilities Provision<Your City Name>
Dataset 33	Area Bifurcation Data :< Your City Name>
Dataset 34	Budget for Open Spaces development :< Your City Name>
Dataset 35	Signalized Intersections :< Your City Name>
Dataset 36	No. of buses :< Your City Name>
Dataset 37	Earnings From bus trips<Your City Name>
Dataset 38	Public Transport Access<Your City Name>
Dataset 39	Public Transport Mode Share<Your City Name>
Dataset 40	Vehicle Registration Date<Your City Name>
Dataset 41	Injuries & Fatalities<Your City Name>
Dataset 42	Water Consumption<Your City Name>
Dataset43	City Consumption<Your City Name>
Dataset 44	Registered Voters<Your City Name>
Dataset 45	Digital Access<Your City Name>
Dataset 46	Diseases<Your City Name>
Dataset 47	Crimes<Your City Name>
Dataset 48	VAT &GST<Your City Name>
Dataset 49	Financial Health :< Your City Name>
Dataset 50	Digital Payments :< Your City Name>

Resource/Dataset Title Nomenclature

Annexure2:NDSAP Assessment Framework Checklist

Policy	Approval of Data City Policy(1.a)	Has the city formally approved the City Data Policy?
	City Data Policy Components (1.b)	Does the City Data Policy have the following section/components: 1.b.i Data Classification 1.b.2 Data Categorization 1:b.3 Data Flow / Approval Framework 1.b. 4 Data Archival and Retention 1.b.s Data Security 1.b.6 Guidelines – Is there any SoP for data collection" 1.b 7 Guidelines - Is there any SoP for electronic data collection? 1.b.8 Guidelines - Is there any SoP for data processing and cleaning? 1.b 9 Guidelines - Is there any SoP for data for quality assessment of data sets? 1.b.10 Guidelines - Is there any SoP for data publishing s per Open Data Norms? 1.b.11 Guidelines - Is there any SoP for engaging stakeholders to assess the data needs? 1. b.12 Guidelines - Is there any SoP for data collection, processing and analysis for on field Survey? 1. b.13 Guidelines - Do the Processes defined include provisions for data analysis?
	Budget for Data initiative(2019-2020)(1.c)	Has the city/municipality earmarked budget in its Annual Budget 2019-20 for data-related initiatives/ activities?
	Budget for Data initiative(2020-2021)(1.d)	Has the city/municipality earmarked budget in its Annual Budget 2020-21 for data-related initiatives/ activities?
People	Appointment of City Data Officer (2.a)	Does your city currently have City Data Officer?
	2.b Appointment of Data Coordinators	What is the percentage of Departments which have appointed Data Coordinators? Total number of departments in which data coordinators have been appointed Number Total number of departments in the ULB
	2.c Data Team	What are the number of members in your data team with their roles and responsibilities? (including Data Scientists, Architect, Analyst, open data expert, interns, outreach experts, and excluding CDO, data coordinators, data champions)
	2.d Capacity Building - Ministry Initiative	How many trainings or workshops on data has the city attended over WebEx to build capacity of its data team for executing the Data Smart Cities Strategy

		from 15th August 2019 onwards?(To be filled by Ministry)
	2.e Capacity Building - City initiative	How many trainings or workshops on data has the city conducted to build capacity of its data team for executing the Data Smart Cities Strategy from 15th August 2019 onwards?
Process	3.a City Data Alliances	How many data-related alliances has the city formed as envisaged in the Data Smart City Strategy?
	3.b Data Hackathons / Data Challenges	Has your city conducted Data Hackathon event/Data Challenge for stakeholders (e.g. Academia, students, research institutes, Start-ups etc.) to help solve city issues through data?
	3.c Solving Urban Challenges using Available Datasets	How many urban challenges (e.g. traffic patterns, safety etc.) has the city identified and attempted to get insights using the available data?
	3.d Analytics capability	Is the city practicing any kind of analytics?
Technology	4.a Sensors for collection of data	Does your city have sensors/field devices that collect data at source?
	4.b Number of datasets	What is the total number of machine readable datasets published by the city on the Open Data Portal?
	4.c Schedule of Updating of Datasets on Open Data Portal	How many datasets has the city updated on the Open Data Portal as per the schedule? Number of datasets uploaded as per the schedule of updation Total number of datasets uploaded
	4.d Number of APIs	How many APIs does the city have for integrating its datasets with the Open Data Portal?
	4.e Spatial readiness	How many city layers (such as roads, water bodies, properties etc.) are mapped on GIS?
Outcomes	5.a Data Stories/Blogs	How many data stories/Blogs has your city published on the Open Data Portal?
	5.b Data related use cases	How many use cases/SoPs of data is the city working on?
	5.c Development of portal/applications	How many services are being delivered through applications on the basis of the city's datasets?
	5.d Alerts & notifications	How many service alerts is the city sending to at least 1% of its citizens? (traffic, disaster, health, water, electricity, environment, etc.)