

# MUZAFFARPUR MUNICIPAL CORPORATION

## City Data Policy

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. This is a draft copy subject to the approval of the Board.

# Contents

<b>1) Abbreviations/Definitions</b>	<b>04</b>
<b>2) Need for Data Policy</b>	<b>06</b>
<b>3) Scope of the Data Policy</b>	<b>07</b>
<b>4) Stakeholders and Collaboration</b>	<b>07</b>
<b>5) Data Management and team structure</b>	<b>09</b>
<b>6) Roles and Responsibilities</b>	<b>10</b>
<b>7) Data Management Principles</b>	<b>12</b>
a. Data categorization	12
b. Data classification	12
c. Data flow/approval framework	13
d. Data archival and retention	13
e. Data security and privacy	14
<b>8) Standard operating Procedures (SOPs)</b>	<b>16</b>
a. SOP for data collection	16
b. SOP for electronic data collection	16
9) SOP for data processing and cleaning	18
10) SOP for quality assessment of datasets	19
11) SOP for data publishing	20
12) SOP for engaging stakeholders	21
13) SOP for data collection, processing and analysis for on-field surveys	21
14) SOP for data analysis	21
15) Open data for different sectors	22

# Message from the Desk of Municipal Commissioner- Muzaffarpur

A city's data is one of its most valuable assets. Urban data is the bedrock of the performance management programs that allow cities to ensure continuous improvement. Reliable data can facilitate interagency collaboration, improve partnerships with the private sector, and expand public engagement. Innovative uses of data allow cities to enforce regulation and improve social services. And, increasingly, open data is serving as the foundation for good government activism, allowing journalists and civic hackers to highlight government inefficiencies or even spot corruption.

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 Muzaffarpur 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.

Shri. Vivek Ranjan Maitrey

Municipal Commissioner

Muzaffarpur Municipal Corporation

# 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/organisation 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.

**MMC** Muzaffarpur Municipal corporation

**MSCL** Muzaffarpur Smart City Ltd

<b>SPV</b>	Special Purpose Vehicle
<b>ICCC</b>	Integrated City Command and Control Centre
<b>ITMS</b>	Intelligent Traffic Management System
<b>CDP</b>	City Data Policy
<b>CDO</b>	City Data Officer
<b>SCDA</b>	Smart City Data Alliance
<b>DCO</b>	Data Coordinator's
<b>API</b>	Application Programming Interface
<b>OGD</b>	Open Government Data
<b>SSL</b>	Secure Sockets Layer
<b>SCADA</b>	Supervisory Control and Data Acquisition
<b>NDSAP</b>	National Data Sharing and Accessibility Policy
<b>NIC</b>	National Informatics Centre
<b>GOI</b>	Government of India
<b>BSPCB</b>	Bihar State Pollution Control Board

DRAFT COPY

# Need for Data Policy

---

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 Muzaffarpur Municipal Corporation as well as entities external to Muzaffarpur Municipal Corporation. This data can be used by various stakeholders for economic, scientific, improve communication, reducing risks, increasing revenues and developmental purposes.

Citizens today expect their cities to deliver robust, user-friendly digital services. Collaboration tools, modern and intuitive websites, mobile applications, self-service portals, and convenient online accounts have become the standard in many facets of life. Expanding digital services in communities make smart cities a more attractive place for residents to live and promote a connected citizen experience.

Accessible government data, interactive maps, government performance dashboards, transparency into budgeting, live-streamed city hall meetings, and a strong social media presence all assist smart cities in creating closer relationships with citizens. These smart technologies help increase civic engagement and trust in city officials.

Urban planning and decision making is a challenging task as it entails complex processes, it involves various actors, and uses data collected from heterogeneous sources. To improve services and to guide in future decision making, all planning decisions are required to be maintained which necessitates the capturing of provenance for smart cities environments. This paper puts forward the idea of tracking the full urban planning process and considers each phase of the process as an individual system.

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 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 Muzaffarpur 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 City Municipal Corporation/Muzaffarpur Smart City Limited, other government departments etc. This document is intended to put in place a formal data governance mechanism at City Municipal Corporation and is an attempt to further enhance the existing data initiatives by having inclusive CDP. Having such a policy in place will make the city of Muzaffarpur being identified as one of those cities that is data-driven and data-sufficient.

# 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 Muzaffarpur; thereby defining the expectations for departmental participation and governance of the data program.

The policy is planned 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 Muzaffarpur Municipal Corporation/Muzaffarpur Smart City Limited. This policy applies to any person/user, organization, administrators, contractors, etc. who intends to access information or assets through any data portal of Muzaffarpur Municipal Corporation/Muzaffarpur Smart City Limited. Specifically, the Data Policy applies to the following information assets of Muzaffarpur Municipal Corporation/Muzaffarpur Smart City Limited:

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

## Stakeholders and Collaboration

---

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

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

- d) **Academia:** Representatives from universities, Colleges, Schools in the city like MIT.
- e) **Policy advocacy groups and NGO's:** 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.

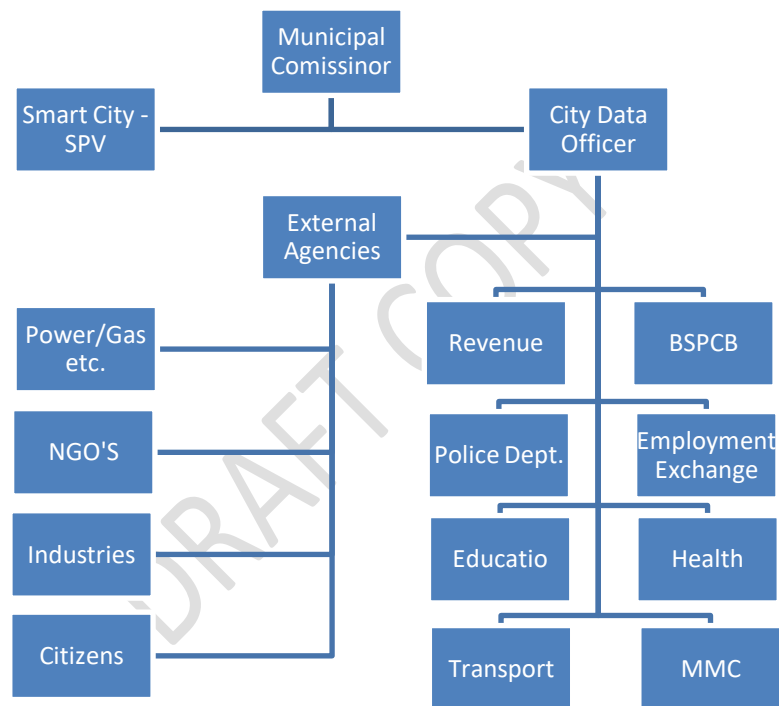
The responsibilities of the CDA will include:

- 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.
- 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 Block chain.
- f) To generate awareness among various stakeholders towards open government initiatives.
- g) To deliver research papers using city data on civic problems in the city.
- h) To organize data-challenges on complex civic problems.
- i) To sensitize ecosystem partners to share data for solving civic challenges.
- j) To support, engage and encourage network/groups/members of data enthusiasts in the city.
- k) To improve city capacity over data driven governance and policy formulation.



# Data Management Team

Muzaffarpur 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 Muzaffarpur Municipal Corporation is as follows:



Further data management team to be extended for developing the data assembly and management if required.

# Roles and Responsibilities

---

Key officials of the data team and their roles and key responsibilities are defined as below:

## **1) City Data officer (CDO)**

The CDO will be the responsible for implementation of the Data Smart Cities Strategy at the city level. CDO will work with city leadership to assess and tap the potential of data and set up data culture across the organization and outside the organization. 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:

- To work as SPOC (Single point of contact) to work as a bridge between City Municipal Corporation, Smart City SPV, Academia, Industry and Citizens to leverage the Data.
- Setup data Teams and data services portfolio. Develop and design data governance program designed to establish data as an asset to be managed, including standards, classification, data and system inventory and data management.
- To chalk out city open data policy in alignment with the Smart City Plan and IT Strategy of City Municipal Corporation.
- Develop policies businesses processes and resources for integrated data sharing across multiple regulatory frame works.
- Responsibilities include management of data collection, data mining, cross channel data integration, predictive analytics and reporting dashboard and data visualization.
- Responsible for collection and aggregation of data and for drawing meaningful insights of it by applying various data analytics tools and technique. To proactively use data driven insights to promote effective urban governance.
- 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.
- Coordinate with MDO (Mission Data Officer) to align with mission data strategy and priorities with respect to open government initiatives and policies.
- Organize regular meetings of the City Data Alliance (CDA)

## **2) Data Champions (DCs)**

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

- Shall identify the datasets/feeds, derived information, intelligence or data challenge with respect to day to day operations of the department.
- 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.

- 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.
- 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:

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

### **4) 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 will include:

- To act as an advisory group to the city leadership on the CDP.
- To assess the data needs of various smart city stakeholders.
- To promote data driven governance and policy formulation.
- To design and implement solutions based on city data.

- To support industry to design solutions using emerging technologies like Artificial Intelligence (AI) and Machine Learning (ML) .
- To assess and design use cases critical to the citizens of the city.
- To generate awareness among various stakeholders towards open government initiatives.
- To bring city's stakeholders on a common platform to influence the city data priorities.
- 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 the city.
- To deliver research papers using city data on civic problems in the city.
- To design and develop prototypes/ solutions on civic problems in the city.
- To organize data-challenges on complex civic problems.
- To organize hackathons and support shortlisted solutions at city level.
- To set up scholarship for postgraduate and graduate students to work with the City Data Team along with the CDO.
- To publish the progress report every month.
- To prioritise the datasets/feeds for publishing on the open data platform.
- To sensitize ecosystem partners to share 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 CDOs by extending resources (like interns, researchers, technology experts), funds (program sponsorship etc.) and technology (solutions etc.).
- To share data available with the partners on data platforms to promote city data.

## Data Management

---

Data Management process will impart a sense of uniformity in the way in which data is generated/ created to the point that it is destroyed/refined/reused. This process will be applicable to all types of data at all levels, all categories and classifications.

### 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 anonymized 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

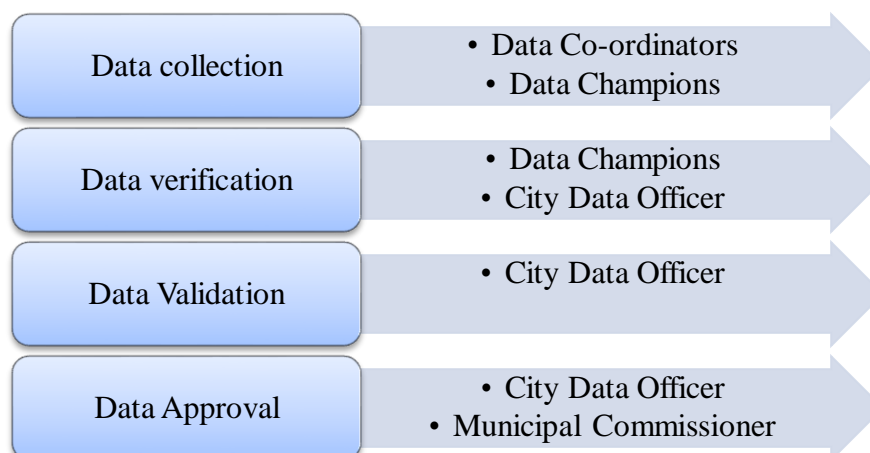
There is a need to classify the data basis its usage and stakeholders. There may be some data which can be open to the public, while some other may be confidential and restricted. Such distinctions need to be appropriately defined to prevent misuse and maintain confidentiality. Muzaffarpur Municipal 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:**

Class	Definition
Public/Shareable Data	Those data not covered under the scope of negative list and non-sensitive in nature. This data is to be provided for public consumption and use.
Negative List	Non-shareable data as declared by the departments/organisations.
Restricted Data	Data which are accessible only through a prescribed process of registration and authorization by respective departments/organizations.
Sensitive Data	Sensitive data as defined in various Acts and rules of the Government of India.

## 3) Data Flow/approval Framework

Muzaffarpur 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 authorize 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 followed by appended below:-



#### **4) Data Archival and Retention**

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

**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 hard-drives, Storages etc subject to the approval of the competent authority.

**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 1 year in a live state so that whenever a report needs to be generated, the data will be extracted from main database. Data older than 1 year will be archived if reporting period extended.

#### **5) Data security and Privacy**

CDO needs 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.

A citizen’s right to privacy emanates from Article 21 on Liberty, as interpreted by the Supreme Court in a judgment. The IT (Amendment) Act, 2008 takes care of privacy rights of consumers by mandating that service providers protect ‘sensitive personal information’. This could include all personal information, financial information such as bank account details, credit card number; biometrics; health information and any other information that is used to identify a person. Data protection new clause 43A: The existing Act provides for penalty for damage to computers, computer systems under the title ‘Penalty and Adjudication’ in section 43 that is widely interpreted as a clause to provide data protection in the country. Unauthorized access to a computer, computer system or computer network is punishable with a compensation of up to one crore rupees. This section has been improved to include stealing of computer source code for which compensation can be claimed. (Computer source has been defined) Data protection has now been made more explicit through insertion of a new clause 43A that provides for compensation to an aggrieved person whose personal data including sensitive personal data may be compromised by a company, during the time it was under processing with the company, for failure to protect such data whether because of negligence in implementing or maintaining reasonable security practices. Further, ‘reasonable security practices and procedures’ will constitute those practices and procedures that protect such information from unauthorized access, damage, use, modification, disclosure or impairment as may be specified in an agreement between the parties or as may be specified in any law in force. In the absence of such an agreement or any law, the central government will prescribe security practices and procedures in consultation with professional bodies or associations.

Data Protection-Security & Privacy Page 8Penalty for breach of confidentiality and privacy: Under section 72 it is presently restricted to those who gain access to an electronic record or document under the powers conferred under this Act. A new section 72A has been added that provides for punishment for disclosure of information in breach of a lawful contract. Any person including an intermediary who has access to any material containing personal information about another person, as part of a lawful contract, discloses it without the consent of the subject person will constitute a breach and attract punishment with imprisonment of up to three years, and/or a fine of five lakh rupees. This is a strong deterrent, and also will bring those responsible for breaching data confidentiality, under lawful contracts, to justice. Along with section 43A, section 72A strengthens the data protection regime in the country. It will go a long way in promoting trusting trans-border data-flows to India. Some of the other laws that have a bearing on data protection and privacy protection are as follows:

- The Indian Penal Code, 1860
- The Indian Telegraph Act, 1885
- The Indian Contract Act, 1872
- The Specific Relief Act, 1963
- The Public Financial Institutions Act, 1983
- The Consumer Protection Act, 1986
- Credit Information Companies (Regulation) Act, 2005Special Legislation(s)
- The Information Technology Act, 2000
- The Information Technology (Amendment) Act, 2008International Conventions
- International Covenant on Civil and Political Rights, 1966
- Universal Declaration of Human Rights, 1948

Apart from the above stated rules some general guidelines to follow for the data security.

- The premises will be physically secured, access control devices should be available for accessing the premises, entry and exit should be monitored.
- Appropriate firewalls, IPS, SSL devices should be used to ensure Network security.
- The solution should support encryption mechanism for transferring data across network and between MMC and server.
- 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 has no longer require access to information systems and services.
- The system should have authentication mechanism either through One Time Password (OTP) or soft token-based technologies for access control and user authentication.
- 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.

- SIEM should be available for Real-time analysis of security alerts generated by applications and infrastructure.
- DAM should be available to monitor all database

All data flows, storage and sharing should adhere to the National and State level data privacy and security policies. 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

DRAFT COPY



# Standard Operating Procedures (SOPs)

---

Standard Operating Procedures (SOPs) are guidelines designed to achieve consistency in specified situations by performing these practices. These SOPs are designed with a view to enhance and standardize data operation and management.

## **1) 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. In order to get the primary datasets we have to collect data from different wards, zones which is under Muzaffarpur Municipal Corporation, Muzaffarpur Smart City Limited employee and Other Government departments. Data which is not going to be used for any kind of analysis or will not be used for any communication purpose should not be collected at all.

If the request is received from external agency:

1. If the request is received from external agency, it 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, the Data Champion/Data Coordinators in consultation with CDO shall take appropriate action.

If the request is received from internal departments:

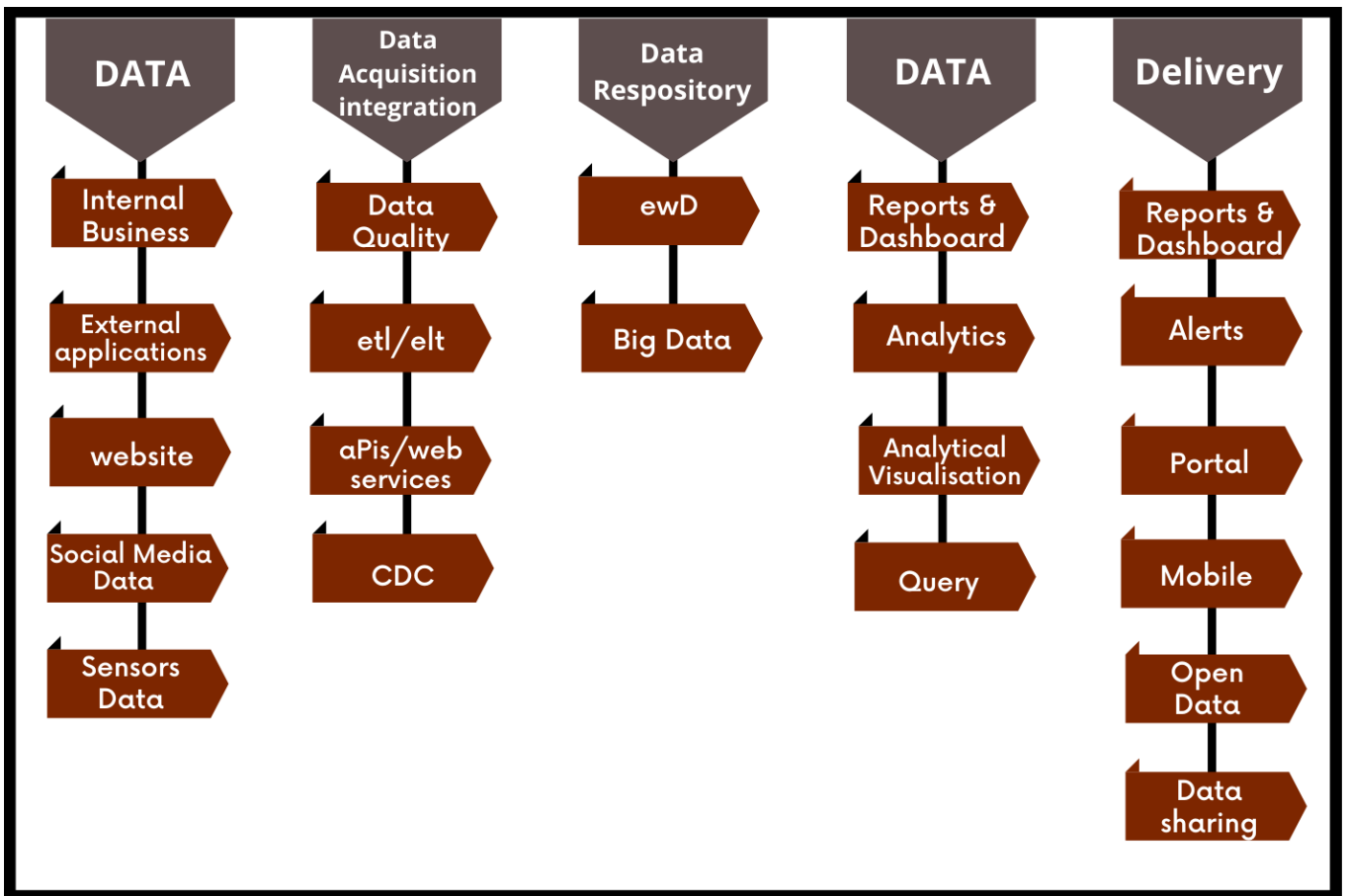
1. If the request is received from internal departments, it should be directed to the Data Champion for the concerned department.
2. The DC of the concerned departments checks for the requested data. If data is available with department, DC shall instruct the Data Coordinators or the concerned personnel to gather the data in requested format.
3. If fresh data is needed to be captured/acquired, the Data Champion/Data Coordinators in consultation with CDO shall take appropriate action.

## **2) 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. The data should be collected with consent of the end-user who may be a citizen or Muzaffarpur Municipal Corporation/Muzaffarpur Smart City Limited employee.
2. Special care needs 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.
3. For data collection and integration with the IoT systems / sensors, the following Data Architecture highlighted below shall be followed by Muzaffarpur Municipal Corporation/Muzaffarpur Smart City Limited.



### 3) SOP FOR DATA PROCESSING AND CLEANING

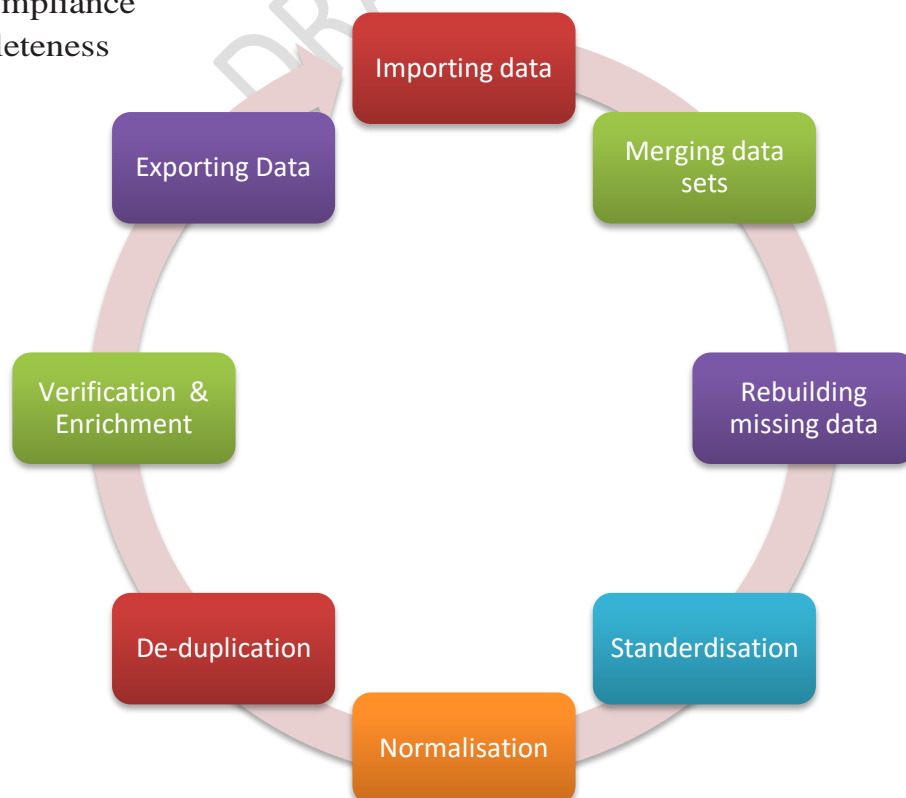
Raw data may be old and inaccurate and can have an adverse impact on results. Data cleaning will be done to ensure that data is correct, consistent and useable by identifying any errors or corruptions in the data, correcting or deleting them, or manually processing them as needed.

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 analysed, 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:

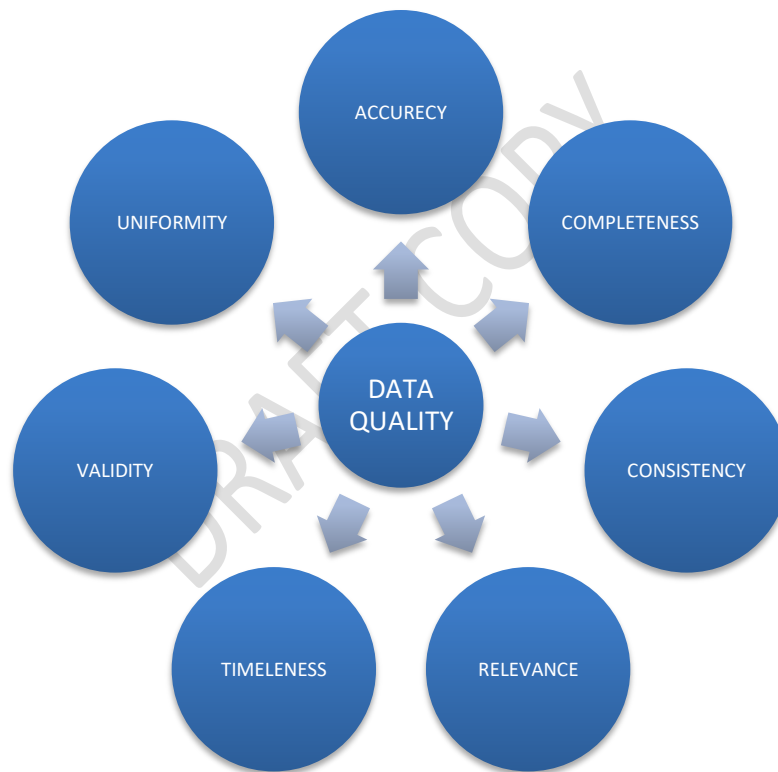
- Spaces in extra columns Compliance
- Ward-wise Compliance
- Blank Cells Compliance
- Standard format Compliance
- All NA Compliance
- Special Characters Compliance
- Split Sheets Compliance
- Datasets Compliance
- Data Completeness



#### 4) SOP FOR QUALITY ASSESSMENT OF DATA SET

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 (Muzaffarpur Municipal Corporation), comprising of CDO, and Additional/Assistant Commissioner may be formed for data quality assessment.



#### 5) SOP FOR DATA PUBLISHING

National Data Sharing and Access Policy (NDSAP) define 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 you have selected 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 finalise deployment:** Obtain the appropriate management approvals for your data based on your selected data publishing option, either from CDO or Municipal Commissioner.

## **6) SOP FOR ENGAGING STAKE HOLDERS**

The concept recognizes the value of enhancing engagement among all four stakeholders of the quadruple- helix model—Government (Muzaffarpur Municipal Corporation Departments, Muzaffarpur Smart City 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
  - › Organising 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.

## **7) SOP FOR DATA COLLECTION, PROCESSING AND ANALYSIS FOR ON-FIELD SURVEY**

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 fulfill a certain gap, with a well-defined problem for investigation. Data collected from on-field surveys can help Muzaffarpur Municipal Corporation/Muzaffarpur Smart City 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.

### 8) SOP FOR DATA ANALYSIS

Data analysis or analytics is an encompassing and multi-dimensional field that uses mathematics, statistics, predictive modeling 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.

Muzaffarpur Municipal Corporation/Muzaffarpur Smart City 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.

### 9) Open Data for Different Sectors

#### Education:-

The application of open data is already changing the ways in some parts of the world, in which teachers teach, students learn, administrators make funding and spending decisions, and graduates find the right employers. The identified five levers that can enable potential value from using open data in education improved instruction, better matching of students to programs, matching students to employment, transparent education financing, and more efficient system administration sources of Open Data in Education Stakeholders.

Sources of Open Data in Education	Stakeholders
<ul style="list-style-type: none"> <li>➤ Public Schools</li> <li>➤ Private Schools</li> <li>➤ Universities</li> <li>➤ Coaching institutions</li> <li>➤ Government Agencies</li> <li>➤ Individuals</li> </ul>	<ul style="list-style-type: none"> <li>➤ Students</li> <li>➤ Families</li> <li>➤ Teachers</li> <li>➤ Administrators</li> </ul>

By which we can improvise the learning plan for the students, real feedback on teaching technology and teachers, professional and personal development along with the technical and analytical knowledge improvement of Students

### Transportation:

Open data will play an important role in improving transportation and believe that it has significant potential to play a role to in addressing the inefficiencies that persist. Many of the improvements we have seen in supply-chain management, logistics, mapping, and route planning over the past two decades were enabled by opening government data from GPS satellites. Transportation customer (defined here as individual travelers as well as organizations and parties that ship cargo) benefit from improved tracking of people and freight. Data streams and data sets on bus and train arrivals, Cargo positions, and aircraft utilization rates are being made open and allow individuals, firms, and government to provide and use services in new ways. There are three major levers for unlocking value with open data in transportation: improved infrastructure planning and management; optimized fleet investment and management; and better-informed customer decision making.

Sources of Open Data in Transportation	Stakeholders
<ul style="list-style-type: none"><li>➤ Government</li><li>➤ Transport Operator</li><li>➤ Individuals</li><li>➤ Third Party data providers</li></ul>	<ul style="list-style-type: none"><li>➤ Transportation Providers</li><li>➤ Airport, Bus Terminus, Railway Station and</li><li>➤ others</li><li>➤ Customers</li><li>➤ Government</li><li>➤ Third Party Developers</li></ul>

### Consumer Products:

Open data not only can accelerate the impact of big data and advanced analytics, but it also can uncover new sources of value for manufacturers, retailers, and consumers in consumer products. Further examining how five levers can deliver value across the consumer products domain through the use of open data:

Sources of Open Data in Consumer Products	Stakeholders
Government Retailers Manufacturers Third Party data Brokers Consumers	Consumers Companies Government

### Electricity:

Open data can help address many of the issues facing the electricity industry, its customers, and government regulators. Making data more open through industry benchmarks can help utilities save on capital investments and operations. As utilities experiment with smart grids and other new technologies, sharing data about technology pilots can avoid duplicated efforts and optimize investment. Sharing data such as technical standards, work processes, task time benchmarks, construction plans, performance management systems, and asset replacement schedules can be used to identify areas for improvement, which can lead to significant savings. Additionally, providing data to electricity users on their usage patterns and showing comparisons with usage by similar customers, can lead to more efficient electricity consumption. While these types of data can help to drive savings, significant investment in technology and operations will be required to fully capture this value.

In the long run, data sharing could help the economic and environmental viability of the energy industry. Lower costs of operations and investment could lead to a more sustainable economic structure, with a larger share of funding derived from current operations and less reliance on future growth to pay for investments. Transparency about the environmental impact per unit of electricity generated can also begin to shift generation investments,

beginning with identifying areas that have the largest potential reduction in pollutants and leading to lower overall environmental impact in the future.

Sources of Open Data in Electricity	Stakeholders
<ul style="list-style-type: none"> <li>➤ Government Agencies</li> <li>➤ Power suppliers</li> <li>➤ Regulators</li> <li>➤ Energy Users</li> <li>➤ Third Party Brokers</li> <li>➤ Industry Groups</li> </ul>	<ul style="list-style-type: none"> <li>➤ Electricity users</li> <li>➤ Government</li> <li>➤ Companies</li> </ul>

**Health Care:**

The use of data to improve the quality of health care and make delivery more efficient is reaching a tipping point, with open and proprietary data playing a leading role. The amount of data being collected, analyzed, and shared among health-care stakeholders has reached critical mass, with growing volumes of digitized medical records, aggregated research and development data, and data that government has collected over the years. These big data sets are yielding critical insights into what are the most effective therapies for specific types of patients, enabling hospitals to isolate common causes of costly hospital readmissions and allowing insurers and other payers to identify variations in care delivery that add needless costs. In addition, individual consumers are finding new ways to manage their health and fitness through My Data, which is made available by care providers or generated by consumers by using exercise monitors and other devices. It can be five levers for capturing value in health care through the use of open and proprietary data. They span the full range of health-care stakeholders from patients to research scientists.

Sources of Open Data in Health Care	Stakeholders
<ul style="list-style-type: none"> <li>➤ Providers</li> <li>➤ Payers</li> <li>➤ Pharma and Medical Product Manufacturers</li> <li>➤ Consumers</li> <li>➤ Government</li> <li>➤ Third Party Brokers</li> </ul>	<ul style="list-style-type: none"> <li>➤ All Stakeholders</li> <li>➤ Payers</li> <li>➤ Pharma and Medical Product Manufacturers</li> <li>➤ Government</li> </ul>

**Real Estate:**

Open data already plays a big role in some areas of residential real estate—using public records to insure titles, for example—but there are many more areas where open data sources have not yet been applied. The five main levers for bringing potential value to developers, home buyers, and lenders: improved location selection for development, improved prioritization of infrastructure investments, optimized seller/leaser and buyer/renter matching, optimized real estate financing, and optimized maintenance planning and execution.