COIMBATORE CITY
MUNICIPAL CORPORATION

COIMBATORE
Smart city

City Data Policy Coimbatore

Contents

Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
CCDA	Coimbatore City Data Alliance
CCMC	Coimbatore City Municipal Corporation
CDA	City Data Alliance
CDO	City Data Officer
CDP	City Data Policy
CSCL	Coimbatore Smart City Limited
CSV	Comma separated values
DC’s	Data Champions
DCO's	Data Coordinators
GML	Geography Markup Language
KML	Keyhole Markup Language
ML	Machine Learning
NDSAP	National Data Sharing and Accessibility Policy
OGD	Open Government Data
PKI	Public Key Infrastructure
RDF	Resources Description Framework
SCADA	Supervisory control and data acquisition
SDG	Sustainable Development Goals
SPV	Special Purpose Vehicles
SSL	Secure Sockets Layer
XLS	Spread sheet - Excel
XML	Extensive Markup Language

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/ organisations.

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.

Chapter 1

1.1 Introduction

In the current age of digital economy, data has proven to be a valuable currency and a crucial asset for improving efficiency, accountability and reliability. The quest to become smarter cities can truly be realized by co-opting the digital revolution unfolding in the country. The importance of data-driven decision making has gradually percolated into the public sector. The importance of open government data for improving transparency and quality of public services is well-established. But the quality of data-centric governance is directly proportional to the maturity of the data ecosystem. The absence of an entrenched data culture impedes innovation, judicious allocation of resources, accountability and citizen engagement.

In order to realize the vision for data-smart cities, it is crucial to establish an institutional and policy framework that clearly articulates the standards and processes for data collection and management in order to improve accountability, avoid misuse and derive discernible decisions for enhancing delivery of public services.

Coimbatore City Municipal Corporation (CCMC)/Coimbatore Smart City Limited (CSCL) aspires to make the data it generates and collects easily accessible to the public for the purposes of empowering its citizens, promoting transparency, innovation and research, institutionalizing data culture and improving the quality of public services. Recognizing the importance of data-driven governance, the City Data Policy (CDP) has been developed for an open, transparent, timely and consistent access to data. The CDP defines and describes the guidelines for collecting and managing the city data and the roles and expectations of government departments in promulgating data-driven governance in Coimbatore.

It is hoped that the guidelines articulated in the city data policy serves as a crucial resource for internal stakeholders (Municipal Commissioner, Smart City CEO, City Data Officer, Data Champions, Data Coordinators, Head of the department etc.) and external stakeholders (civil society, academia, private sector etc.) to actively collaborate with the city on data related outreach and programs.

1.2 Need for Data Policy

Cities are vibrant ecosystems characterized by multitude of public and private stakeholders, who in the process of performing their integral roles, generate troves of data. However, the available data remains in silos in the absence of overarching guidelines and protocols on reliable data management. Which often results in duplication of efforts, misallocation of resources and strained citizen engagement. In order to unlock the true potential of urban data, a definitive set of guidelines is required to define the data ecosystem.

The CDP of CCMC/CSCL envisages to define the contours of the data landscape including protocols on data generation, data standards, data sharing, data ownership, data privacy etc. It aims to provide an integrated platform for easing access to open data generated by various departments of CCMC/CSCL.

1.3 Scope of Data Policy

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

a) Data/information collected, captured, aggregated, processed and shared by CCMC/CSCL
b) Citizens data/information
c) Personnel data/information relating to employees of CCMC/CSCL

1.4 Data Policy Goals for CCMC

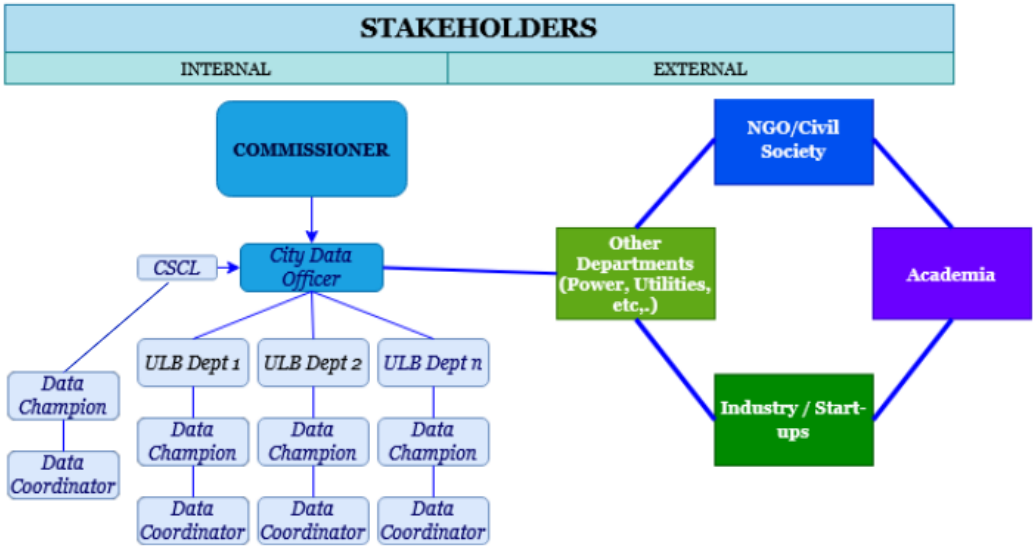
- Institutionalization of data culture
- Citizen Engagement and Empowerment
- Multi-disciplinary research and innovation
- Transparency and Accountability
- Data Sharing Exchanges

1.5 Data Policy Pillars

Chapter 2

2.1 Data Management Team

CCMC/CSCL will set up a City Data Committee for management of data at the city level and enabling coordination between various departments and external stakeholders for making sure that the data is available as and when required. CCMC/CSCL will also engage and secure buy-in from both internal and external stakeholders on key decisions. Further, this will help navigate through complicated hurdles (e.g. bureaucratic, political etc.) and to take prompt decisions and actions pertaining to collection, segregation and release of open data.



2.2 City Data Officer

The City Data Officer (CDO) will be the officer responsible for implementation of the Data smart Cities strategy at the city level. CDO will assess and tap the potential of data and set up data culture across the organisation and outside the organisation. CDO will act as single point of contact to all internal and external stakeholders in the city.

Responsibilities:

a) The CDO will develop a CDP which will be reviewed every month to keep it contextual to the need of the times.
b) Coordinate with Mission Data Officer (MDO) to align with mission data strategy and priorities with respect to open government initiatives and policies.
c) Organise regular meetings of City Data Alliance.
d) Coordinate with officers of various other government departments/agencies within the city for the effective implementation of City Data Policy.
e) Assess all proposed or under implementation projects to identify the datasets/feeds which could generate public datasets/feeds or may be useful for internal analysis.

f) Assess all periodic and recurring MIS needs to identify the datasets/feeds which could be shared with other departments through data exchange.
g) Assess all the operational IT Projects for identifying public datasets/feeds.
h) CDO will publish data Catalogues and Data Sets/Feeds on Open Government Data (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.

2.3 Data Champion

Data champions (DCs) 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.

Responsibilities

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

2.4 Data Coordinators

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

Responsibilities

a) Aggregate the data demand from various channels.
b) Sensitizing the department employees over the importance of data quality etc.
c) Perform collection, interpretation and recording of data in accordance with CDP standards and CDO guidelines.
d) Perform data validation and ensure data quality.
e) Sort and organise the data; both hard copy and electronic versions.
f) Transmit data report to CCMC/CSCL or CDO via Internet.
g) Update CCMC/CSCL website or Coimbatore OGD Portal with latest data records.
h) Assist department staff in data entry when required.
i) Provide data management updates in all internal and external meetings as required.
j) Analyse data for quality improvement purposes.

k) Prepare data for reporting, meetings and presentations for the concerned department and CCMC/CSCL at large.
l) Ensure data management procedures comply with CDP.
m) Provide statistical analysis and longitudinal analysis of data.
n) Prepare and submit data required for audits.

2.5 Coimbatore City Data Alliance

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

The CCDA will provide a collaborative framework to create and define use cases to solve critical city problems through the use of data, catalyse 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.

Roles of Coimbatore City Data Alliance (CCDA):

a) To act as an advisory group to the city leadership on the CDP.
b) To promote data driven governance and policy formulation.
c) To design and implement solutions and analysis using city data.
d) To support industry to design solutions using emerging technologies like Artificial Intelligence (AI), Machine Learning (ML) and Blockchain.
e) To assess and design use cases critical to the citizens of the city.
f) To generate awareness in various stakeholders towards open government initiatives.
g) To facilitate data for co-creation and collaboration over civic issues
h) To provide critical feedback to the city over the quality and relevance of data provided by city.
i) To deliver two research papers annually using city data on civic problems in the city
j) To design and develop two prototype/ solutions annually on civic problems in the city
k) To organize a data-challenge every half year on complex civic problems
l) To organize a Hackathon annually and support shortlisted solutions at city level
m) To set up scholarship for postgraduate and graduate interns to work with the city data team along with the CDO.
n) To publish the progress report every month
o) To prioritize the data Sets/feeds for publishing on data platform
p) To sensitize ecosystem partners to share the data for leveraging data for solving civic challenges
q) To support, engage and encourage network/groups/members of data enthusiasts in the city
r) To improve city capacity over data driven governance and policy formulation

s) To support CDO by extending resources (like interns, researchers, technology experts), funds (program sponsorship etc.) and technology (solutions etc.)
t) To share data available with partners on data platform to promote city data

2.6 Stakeholder for Coimbatore City Data Alliance

The following stakeholders are envisaged as the key constituents of the CCDA. The CCDA will assess, strategize, plan, implement and review the CDP:

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

Chapter 3

3.1. Components of Data Policy

Implementing the data policy requires an extensive process that integrates all the facets of data management. Having such a process in place will institutionalize a sense of uniformity in the way in which data is generated/created, used and shared. The following process will be applicable to all types of data at all levels, all categories and classifications.

3.2 Data Categorization

Data will be categorized into two broad categories:

1. Personal Data: Personal data is that data which is specific to a particular individual. It is the responsibility of the CCMC/CSCL that they do not, in any case, publish personal identifiable data/information or parts of personal data/information on any of the Open Data Platforms or Datasets. It becomes the responsibility of the CDO to ensure that all personal data is anonymised before it is published.

2. 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.

Personal and Non-Personal Data will be classified into following category:

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

3.3 Data Classification

Different types of datasets generated both in geospatial and non-spatial form by different ministries /departments 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. CCMC/CSCL will prepare the negative list within one year of the notification of the policy, which will be periodically reviewed by the assessment committee.

Classification of Datasets

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:
    I. **Title (Required)**: A unique name for the catalogue (group of resources) viz. Current Population Survey, Consumer Price Index, Variety-wise Daily Market Prices Data, State-wise Construction of Deep Tube wells over the years, etc.
    II. **Description (Required):** A detailed description of the catalogue e.g., an abstract determining the nature and purpose of the catalogue.
    III. **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.
    IV. **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.
    V. **Sector & Sub-Sector (Required):** The sectors(s)/sub-sector(s) those most closely apply(ies) to your catalogue.
    VI. **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 specified under 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:** As per NDSAP following data formats should be published:

    1. CSV (Comma separated values)
    2. XLS (Spread sheet - Excel)
    3. ODS (Open Document Formats for Spreadsheets)
    4. XML (Extensive Markup Language)
    5. RDF (Resources Description Framework)
    6. KML (Keyhole Markup Language used for Maps)
    7. GML (Geography Markup Language)
    8. 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 Facebook, Twitter.

i) **Ownership**: All data sets/ feeds remain property of CCMC/CSCL. The CDO will endorse Government Open Data License to ensure that published data is not misused or misinterpreted by its users.

3.4 Data Archival and Retention

Data Retention

CDO to undertake retention of data, i.e. the maintenance of documents to be accessed by an authorised user at a later stage. Retrieval schedule of the data will be as per the rules and regulations defined by the Government of India:

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

Data Archival

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

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

1) Category-I (e-Files/records to preserved permanently which are of historical importance) – For 5 years, it will be kept in the Department's server and thereafter transferred to other available physical storage formats such as Tapes, hard-drives, Storages etc.

2) Category –II (e-Files/records of secondary importance and have a reference value for a limited period) – 5 years on the Department's server. In exceptional cases, if the 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 state so that whenever a report needs to be generated, the data will be extracted from main database. Data older than 6 months will be archived. If report duration extends beyond 6 months, the data will be retrieved from archival to generate the report.

3.5 Data Security and Privacy

CCMC/CSCL will manage the data security and privacy and ensure that data is protected from loss, unauthorised use and corruption. All data flows, storage and sharing should adhere to the National and State level data privacy and security policies already in place.

Data security and privacy will be handled under the following operating principles:

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

The data management by CCMC/CSCL will have the following security features:

a) Data dissemination should be only to authenticated and authorised stakeholders (both internal and external) through data fiduciaries.
b) The usage rules for data elements must specify for what purposes the data can or cannot be used. The solution should support Secure Sockets Layer (SSL) encryption mechanism for transferring data across network.
c) The data transferred across network should be encrypted using Public Key Infrastructure (PKI).
d) Access to all system resources including data files, devices, processes and audit files should be provided to the intended users only.
e) All mobile applications should be designed and developed in a way that it ensures security of the application and data on the device.
f) Documents should be assigned 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) Database Activity Monitoring (DAM) should be available to monitor all databases.

3.6 Data Flow

CCMC/CSCL will set up enterprise processes to control the existing available data within the City administration. At every stage of data generation, the concerned stakeholders shall approve and authorise the data usage.

Data flows may vary according to different scenarios, such as data being circulated between departments, uploading data on the open data portal, sharing data with third party, etc. Each dataset has a trustee accountable for data quality and security. Appropriate data flow and approval mechanisms should be in place for such situations.

Data Flow and Approval

Chapter 4

4.1 Standard Operating Procedures (SOP)

Standard Operating Procedures (SOPs) are succinct guidelines designed to achieve consistency in specified situations by postulating a standard practice in performing those functions. These SOPs are designed with a view to enhance and standardise data operation and management.

4.2 SoP for Data collection

If the request is received from external agency:

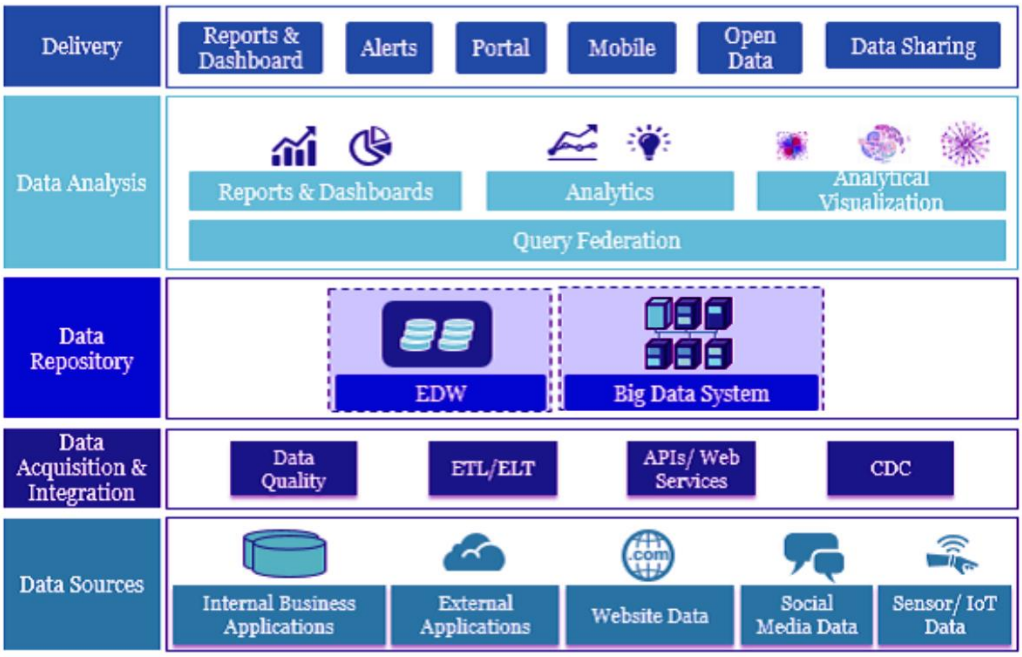
a) If the request is received from external agency, it should be directed to the CDO.
b) Depending on the requested data or the data which need to be collected, the CDO shall direct the request to Data Champion (DC) of the concerned department.
c) The DC of the concerned departments checks for the requested data. If data is available with department, DC shall instruct the Data Coordinators or the concerned personnel to gather the data in requested format.
d) DC will take approval of data from their HOD.
e) If fresh data is needed to be captured/acquired, the Data Champion/Data Coordinators in consultation with CDO shall take appropriate action.

If the request is received from internal departments:

a) If the request is received from internal departments, it should be directed to the Data Champion for the concerned department.
b) 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.
c) If fresh data is needed to be captured/acquired, the Data Champion/Data Coordinators in consultation with CDO shall take appropriate action.

4.3 SoP for Electronic data collection

a) Data which does not fall under the use for analysis or communication should not be gathered
b) Any data collected should be upon written consent from the end-user who may be a citizen, organization or CCMC/CSCL official
c) Data security guidelines mentioned under the “Data Security” section should be followed to maintain the privacy and security of the data.
d) The Data Architecture explained below should be followed by the CCMC/CSCL officials for data collection and integration with sensors and IoT systems.



The CCMC/CSCL data architecture is explained below:

Data Sources: It 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. CCMC/CSCL could leverage various unstructured data emerging from different sources and third-party systems like social media, internet, websites, videos, images etc.

Data Collection and Acquisition: Data Collection and Acquisition Layer acts 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.

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.

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.

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

Life Cycle of Data

a) Field Data: CCMC/CSCL should deploy various IoT sensors, actuators, devices, cameras and solutions to capture the data from the field directly.
b) Operations Data: CCMC/CSCL should deploy various system under various 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. These systems generate various data points in different formats which provides critical information and intelligence to officers to manage critical services and Infrastructure of the city. There should be arrangements to collect, store, integrate, analyze and generate meaningful insights from this data.
c) Third Party Platform and Mobile Apps: Citizens also avail services from private businesses like radio taxi, food delivery, hospitals and labs etc. which could provide various insights and patterns which could be instrumental in policy formulation and city planning.
d) Internet: Various platforms engages citizens and communities which captures general sentiments which could provide insights mood or opinion of citizens and communities towards specific issue.

These data sources and systems could generate structured and unstructured data.

a) Structured Data: Data which is generated by systems or humans and could be handled using existing or predefined models. Structured data could be stored in relational data bases and analyzed using basic search algorithms. E.g.: Location data, User transactions, Sensor data etc.
b) Un-Structured Data: Data which is generated by systems or humans and cannot be handled using existing or predefined models. Un-structured data cannot be stored in relational data bases and is difficult to analyze using basic search algorithms. E.g.: Images, Video files, Audio Files etc.

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 formats generated from various systems which are as follows:

a) Field Sensors and Devices:

- Hierarchical files (JSON, XML, YAML, etc.)
- Real-time stream
- Objects
- Videos
- Images
- Locational Data

b) Operations Systems

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

c) Internet and Social Media Content:

- Blogs
- Video
- Music
- Sentiments
- Images

Under CDP, 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. It's often used to build a data warehouse. During this process, data is taken (extracted) from a source system, converted (transformed) into a format that can be analyzed, and stored (loaded) into a data warehouse or other system.

Data Analysis

Analytics is an encompassing and multidimensional field that uses mathematics, statistics, predictive modeling and machine-learning techniques to find meaningful patterns and knowledge in recorded data. Applying intelligent techniques to uncover insight from the relevant data. CCMC/CSCL to set up Analytics Division to accomplish Data Analysis insights from the Data sourced and integrated.

Few approaches include:

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.

Data Intelligence

It is final stage of the journey from being informed to actionable insights and finally to actions. The Analytics team of CCMC/CSCL under the leadership of CDO shall perform (but 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.

4.4 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.

a) 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.
b) If there is existing data, identify discrepancies which may come from different sources.
c) The collected data shall be properly processed and cleaned before performing any kind of analysis.
d) If needed, commercial software available in the market can be used with prior approvals from the concerned authorities.

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

4.5 SoP for Quality assessment of datasets

Quality assessment of data is needed to ensure that the quality standard is met i.e. accuracy, free from any sort of legal issues, privacy of an individual is maintained and does not compromise with the National security. Data quality shall be assessed from the perspectives of adequacy, appropriateness, accuracy and reliability, authenticity, consistency and validity.

a) An assessment committee comprising of the Deputy Commissioner, Assistant Commissioner Accounts, Chief Health Officer, Executive Engineer, Water Supply and Executive Engineer, Solid Waste Management will be constituted under the leadership of the Smart City CEO. External members can be nominated to the committee on a case to case basis.
b) The City Data Officer and Data coordinators from the respective departments shall submit the data to the assessment committee for data quality assessment.

4.6 SoP for Data publishing as per Open Data Norms

a) Only data which has been approved by Assessment committee and Municipal Commissioner shall be uploaded on Open Data Portal.
b) Data sets which are considered to be open by default unless classified as internal, sensitive, protected or restricted shall be uploaded on the open data portal.
c) Data sets and feeds must be published with proper metadata. Information about the datasets being published using common data taxonomy/structure shall be uploaded as it helps in providing easy access through Data Platform.
d) Data Sets and feeds should be published in formats specified under NDSAP open format.
e) Data should be provided in freely available formats which can be accessed without the need for a software license.
f) Data sets and Feeds should be machine readable.
g) It is proposed that data should be published in any of the following formats:

    1. CSV (Comma separated values)
    2. XLS (Spread sheet - Excel)
    3. ODS (Open Document Formats for Spreadsheets)
    4. XML (Extensive Markup Language)
    5. RDF (Resources Description Framework)
    6. KML (Keyhole Markup Language used for Maps)
    7. GML (Geography Markup Language)

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

4.7 SoP for Engaging stakeholders to assess the data needs

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

Key activities may include:

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

4.8 SoP for Data collection, processing and analysis for on-field survey

Depending on requirement of data, competent agency can be employed to conduct field survey. Going-forward, all the e-governance IT applications/ Systems shall be designed in such a way that manual processes get replaced by automated process without much human intervention. As most of the processes would be automated and handled by electronic mode, data will be readily available for further analysis

4.9 SoP for Data Monetization

CSCL will be in charge of identifying datasets which can be monetized. Further, CSCL will develop a data monetizing platform in order to channelize revenue sources that would make the data ecosystem self-sustainable. The datasets will be differentially priced based on various criteria's like type of stakeholder, type of usage etc. The pricing of the datasets will be determined and fixed by the CSCL Board of Directors and approved by the Municipal Administration and Water Supply (MAWS) department at the state level.

Chapter 5

5.1 Data Policy Budget

In order to handle the complexities of the data, it is important to judiciously allocate resources at critical points in the data life cycle. The availability of the resources will be crucial for avoiding duplication, incorrect data entry, cleansing techniques and holistic civic engagement.

The budget for implementing the CDP will be determined by the Board of Directors of CSCL. The personnel requirements will be sourced from the TULIP program. The associated costs for capacity building and institutional strengthening will be borne by CSCL.

Components	Indicators	
Data Cell Personnel	Data Scientist	1 key person
	Data Analyst	2 resources
	Data Operators	5 resources
	Data Enumerators	5 resources
Capacity Building	Training and workshops	At least four training workshops for city officials per year
	Data Hackathon/Challenges	At least one Open Data related challenges per year
Infrastructure	Development, Operation and Maintenance of IT Infrastructure and security systems	
	Development of Web Portal	
	Data Collection infrastructure	
Citizen Engagement	Data Town halls	6 sessions per year (bi-monthly)
	Publication of City Data, Data Stories and Case studies	
	Identification of priority datasets	
	Public feedback and citizen generated data	
	Engagement of Academia, NGOs and Private sector for developing innovative products using available data	

Amendments

Technology and database ecosystem are a highly dynamic and constantly evolving field. Hence, the innovative potentials, needs and challenges of data-centric governance keep on evolving with time. In order to keep this data policy relevant and appropriate with time, the policy will be reviewed annually by the City Data committee in consultation with the relevant experts and stakeholders. The committee would be empowered to make necessary amends to the policy.

Conclusion

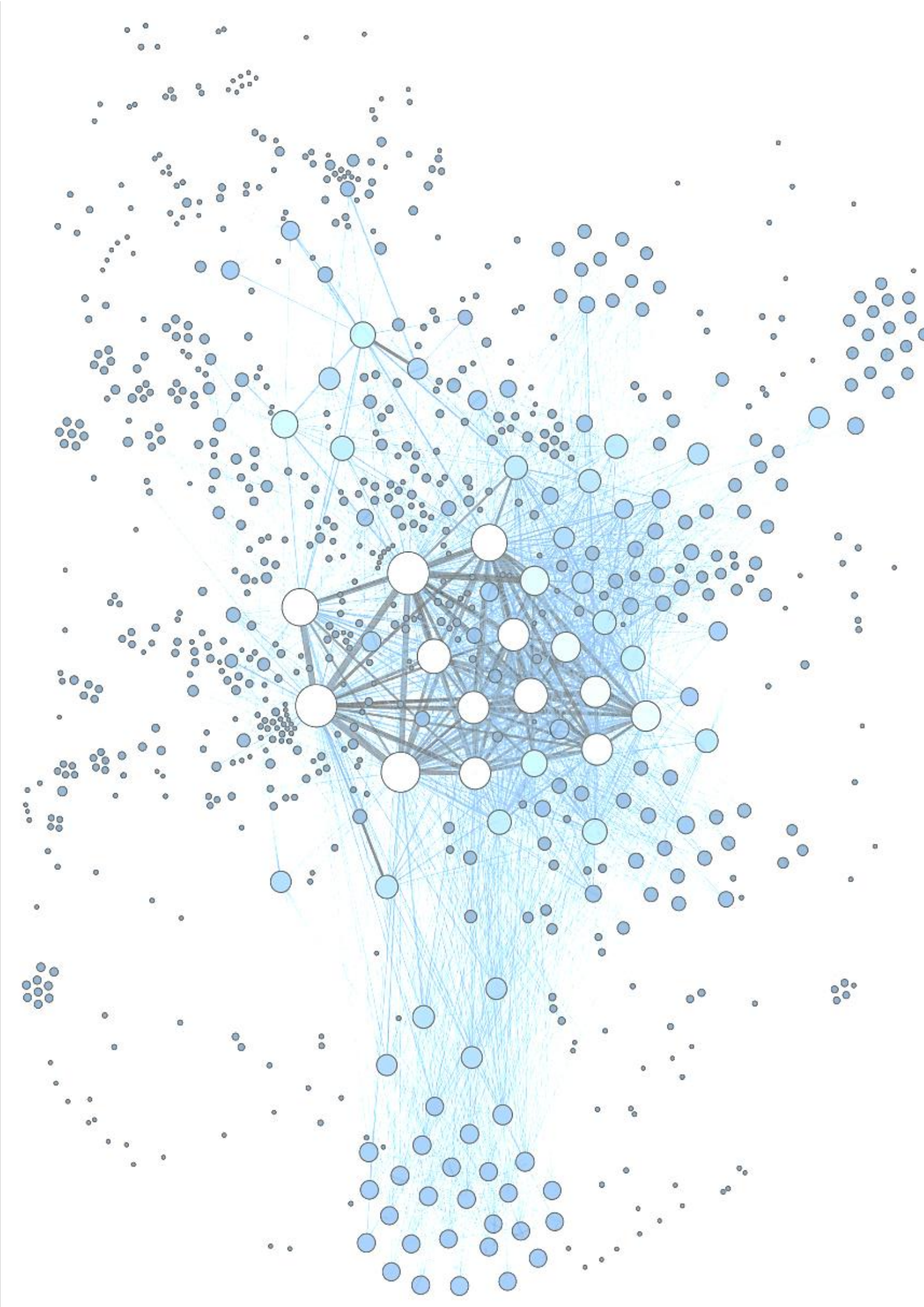
Data and digital technologies have fundamentally transformed the approach to development of the urban ecosystem. Effective utilization of data can aid in identifying the most pressing issues in the city while also offering innovative avenues to solve these problems.

The City Data Policy for Coimbatore aims to promote a technology-based culture of data management as well as data sharing and access. It proactively opens up the data, which could be further utilised by public and private stakeholders for developmental purposes and solving urban challenges. The guidelines articulated in the CDP is well positioned to improve the city's capacity to generate and manage data and, in the process, improve the quality of the city's reporting on various urban missions, Sustainable Development Goals (SDG) etc., and forge effective and innovative data partnerships with private sector, academia and civil society.

Annexure

Data Maturity Assessment Framework Checklist

Component	Indicator
Policy	Approval of City Data Policy
	Components of City Data Policy
	Allocation of City Budgets
	Approval of Budget
People	Appointment of City Data Officer
	Appointment of Data Co-ordinators
	Roles and Responsibilities of Data Co-ordinators
	Additional Team Members for Data Initiatives and
	Trainings and Workshops to Build Capacity
	Percentage of Officials Trained to Implement DataSmart Cities Strategy
Process	Formation of City Data Alliance
	Innovation Hackathon Events/Co-Creation Challenges/ Data Challenges for Academia and Students
	Solving Urban Challenges using Available Datasets
	Identification of additional datasets for answering policy questions
	Access to training infrastructure
Technology	Active Open Data Portal
	Electronic Collection of Data
	Machine Readable Data Sets on OGD Portal
	Schedule of Updation of Datasets on OGD Portal
	Compliance with Schedule of Updation of Data Sets on OGD Portal
	Availability of Sensors/Field Devices to capture
	Digitization of Citizen Centric Services
Outcomes	Positive Citizen Engagement Outcomes
	Development of Applications on the City ODP
	Data related use cases
	Key Urban Challenges Addressed



Amendments

Technology and database ecosystem are a highly dynamic and constantly evolving field. Hence, theinnovative potentials, needs and challenges of data-centric governance keep on evolving with time. In order to keep this data policy relevant and appropriate with time, the policy will be reviewed annually by the City Data committee in consultation with the relevant experts and stakeholders. The committee would be empowered to make necessary amends to the policy.

Conclusion

Data and digital technologies have fundamentally transformed the approach to development of the urbanecosystem. Effective utilization of data can aid in identifying the most pressing issues in the city while also offering innovative avenues to solve these problems.

The City Data Policy for Coimbatore aims to promote a technology-based culture of data management as well as datasharing and access. It proactively opens up the data, which could be further utilisedby public and private stakeholders for developmental purposes and solving urban challenges. The guidelines articulated in the CDP is well positioned to improve the city's capacity to generate and manage data and, in the process, improve the quality of the city's reporting on various urban missions, Sustainable Development Goals (SDG) etc., and forge effective and innovative data partnerships with private sector, academia and civil society.

28/12/2021
CDO

Commissioner,
Coimbatore Corporation,
Coimbatore.

DEPUTY COMMISSIONER,
CITY MUNICIPAL CORPORATION,
COIMBATORE - 1.