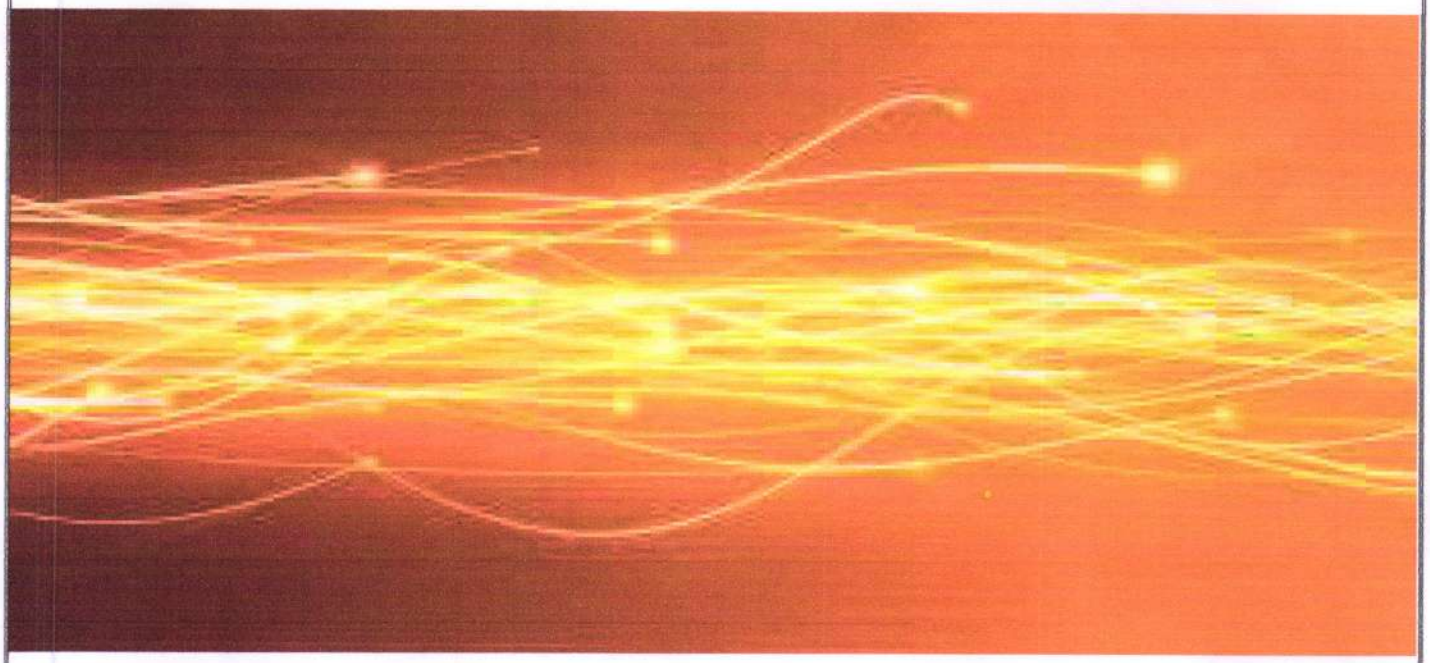
पुणे स्मार्ट सिटी डेव्हलपमेंट कॉर्पोरेशन लिमिटेड
PUNE SMART CITY DEVELOPMENT CORPORATION LTD.

City Open Data Policy

For

Pune City



Document Control	
Document Name	City Data Policy for Pune
Prepared By	PMC IT, Department
Version No	1.1
Classification	Public
Date	December 16,2020

Contents

1. **Abbreviations and/or Definitions** 7
2. **Need for Open Data Policy for Pune** 8
3. **Pune Open Data Policy and Scope** 9
   - 3.1. Purpose 9
   - 3.2. Open Data Policy Goals 9
4. **Data Lifecycle Management** 9
   - 4.1. Data Standards 11
   - 4.2. Data Categorization 12
   - 4.3. Data Classification 12
   - 4.4. Data Security and Privacy 13
   - 4.5. Updating and Maintenance of Data Sets 13
5. **Data Set Provisioning and Identifying Data Sets** 15
6. **Stakeholders and Collaboration** 16
7. **Data Management and Team Structure** 17
8. **Roles and Responsibilities** 18
   - 8.1. City Data Officer (CDO) 18
   - 8.2. Open Data Champions 19
   - 8.3. Open Data Coordinators 19
9. **Open Data Policy – Implementation Plan** 21
10. **Standard Operating Procedures** 22
    - 10.1. SoP for Data Collection 22
    - 10.2. SoP for Electronic Data Collection 23
    - 10.3. SoP for Data Processing and cleaning 24
    - 10.4. SoP for Data for Quality Assessment and Data Sets 24
    - 10.5. SoP for Data Publishing as per Open Data Norms 25
    - 10.6. SoP for Engaging Stakeholders to Assess the Data Needs 26
    - 10.7. SoP for Data Collection, Processing and analysis for on field survey 27
    - 10.8. Data Analysis for collected data 28

Annexures 29

Message from the Desk of Hon'ble Additional Commissioner of Pune Municipal Corporation

The 21st Century is the century of data. The continuous and increased demand for data defines who we are, what we do, where we go and how we get there. Recognizing that timely and consistent access to public data is an essential component of an open, transparent, collaborative and effective government, Pune City Government is committed to proactively sharing its data as broadly as possible via an Open Data Portal. To achieve the same, Pune Municipal Corporation has developed a dedicated Open Data Platform for the Citizens of Pune – http://opendata.punecorporation.org. Now the portal is developed, it's the need of time to design and implement a City Data Policy at city level for the effective communication and coordination between the citizens and government.

The main aim for designing this open data policy is to ensure that data and information are easily accessible by the public and it is within the legal and legislative framework defined by Pune City. Also, this document will cover the importance and need of City Data Policy at city level.

As the Additional Commissioner of Pune Municipal Corporation of Pune City, I congratulate team PMC and ensure full support for the implementation of City Data Policy at city level.

Shri. Dr. Kunal Khemnar, IAS
Additional Municipal Commissioner
Pune Municipal Corporation

Message from the Desk of IT HoD, Pune Municipal Corporation

With the emergence of data-driven innovation hubs and smarter cities across India, the potential of data has become beyond dispute. Considering this fact, Open Data Policy has been considered as one of the most important public policy and the potential impacts of sharing such data cooperatively are enormous. Its impacts are already visible in our societies today, but there remains a need for further research in order to pinpoint the exact benefits of such information transformations and how they affect us as individuals.

Pune city has already developed a dedicated open data portal. Due to this reason, Pune city has chosen Open Data as one of its focus point. As a next step, Open Data Policy will help the city to foster Open Governments across Pune City and a better participation of citizens in the decision-making processes of Pune City, to increase transparency of public spending and political handling.

For the effective and timely implementation of this policy, I urge the city and all 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. Rahul Jagtap
HoD – Department of Information Technology
Pune Municipal Corporation

Sr. No.	Policy Name	Details
		Open Government Data Platform India may request the user to provide personal information, he/she will be informed for the particular purposes for which the information is gathered and adequate security measures will be taken to protect user's personal information.
4.	**Copyright Policy**	Material featured on this Portal may be reproduced free of charge after taking proper permission by sending a mail on OGD platform website. However, the material has to be reproduced accurately and not to be used in a derogatory manner or in a misleading context. These terms and conditions shall be governed by and construed in accordance with the Indian Laws.
5.	**Content moderation and approval policy**	India's Open Government Data Platform – data.gov.in is a repository of large number of datasets meticulously divided into catalogues. All the datasets available are made available in adherence with the principles of Open Data Movement, which mandate easy access for further reuse in machine readable format. This particular platform is basically provides its user with important datasets of the various government departments without comprising its original value. The moderation process here is confined to formatting and ensuring that the datasets are made available in concurrence with the principles of Open Government Data, where no compromise or addition is made to the original values of the datasets which are provided by the data owners. The process of making the datasets into Open format forms the base for standardization in this particular platform.
6.	**Content Review Policy**	Open Government Data Platform – data.gov.in has a very robust policy of review its content through routine testing of web pages and the content. Since this is a platform which boasts of datasets and catalogue, the Project Management Unit (PMU) is frequently in the process of updating and reviewing the datasets and catalogues in order to provide up to date datasets to its users.

Sr. No.	Policy Name	Details
7.	**Content Archival Policy (CAP)**	As per Open Government Data Platform archival policy, nothing is archives right now. All the contents contributed by various stakeholders are accessible on the portal.
8.	**Blog Disclaimer**	The opinions expressed in blogs are the personal opinions of the authors/users.

2. Standardization of Data Access as per National Data Sharing and Access Policy (NDSAP)

National Data Sharing and Access Policy defines standards for publishing data sets and feeds. These standards need to be adhered to by Pune City while sharing its data sets.

- **Open by Default:** Datasets are considered to be open by default unless classified as internal, sensitive, protected or restricted.
- **Meta Data:** Datasets 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.
- **Data Catalogue:** As per NDSAP metadata elements for data sets or feeds are defined as follows:

  a. **Title (Required):** A unique name for the catalog (a 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.

  b. **Description (Required):** Provide a detailed description of the catalog e.g., an abstract determining the nature and purpose of the catalogue.

  c. **Keywords (Required):** It is a list of terms, separated by commas, describing and indicating at the content of the catalog. Example: rainfall, weather, monthly statistics.

  d. **Group Name (Optional):** This is an optional field to provide a Group Name to multiple catalogues to show that they may be presented as a group or a set.

  e. **Sector and Subsector (Required):** Choose the sector (s)/subsector (s) that most closely apply to your catalogue.

  f. **Asset Jurisdiction (Required):** This is a required field to identify the exact location or area to which the catalogue and resources (Dataset/Application) cater to namely entire country, state, district, city etc.

- **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.
- **Machine readable:** Data sets and feeds should be machine readable
- **Formats:** As per NDSAP following data formats should be published:

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

References:

1. https://data.seattle.gov/stories/s/Data-Policy/6ukr-wvup/
2. http://www.minneapolismn.gov/policies/opendata
3. https://www.nlc.org/sites/default/files/2017-03/CSAR%20Open%20Data%20Report%20FINAL.pdf

1. Abbreviations and/or Definitions

Terms	Description
Data	Unprocessed or raw information that is collected for a prescribed business function
Information	A collection of data or documents that is processed, analyzed, interpreted, organized, classified or communicated in order to serve a useful purpose, present facts or represent knowledge in any medium or form
Data Archive	A place where machine-readable data are acquired, manipulated, documented, and distributed to others for further analysis and consumption
Data Generation	Initial generation / collection of data or subsequent addition of data to the same specification.
Open Data	Defined data that is freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control.
Dataset	Collection of related set of information that is composed of separate elements
Geospatial Data	All data which is geographically referenced
Metadata	Data about data, Metadata informs the user of who, when, what, where, why, and how data were generated. Metadata allows the data to be traced to a known origin and know its quality.
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
Sharable data	Those data not covered under the scope of negative list and non-sensitive in nature
Pune City Data	All data created, collected and/or maintained by the Pune City
Open Data Portal	The platform created and maintained by Pune Municipal Corporation for defined datasets http://opendata.punecorporation.org/Citizen/User
City Data Cell	A city level Analytics and Management Unit
NDSAP	National Data Sharing and Accessibility Policy designed and published by Government of India
OGD	Open Government Data

2. Need for Open Data Policy for Pune

Pune Municipal Corporations is one of those leading city administrations in India that has taken huge strides in inculcating a Digital and an IT Empowered eco-system in its quest towards becoming a Smart City. To name a few of its accomplishments Pune is one of India's first City to have migrated it's IT applications including its Websites, Portals, Applications and Software on the Cloud environment. PMC, Pune's Civic authority, has dedicated Web portals and Mobile applications to address grievances of citizens. The corporation has successfully implemented its GIS portal with various layers that plot the entire city of Pune on a digital map across various layer categories. Apart from Digital and IT initiatives, PMC is also aggressively making newer developments in the field of Health, Education, Civic Infrastructure, Water Projects and other major civic initiatives.

Initiatives like these and various other projects are a fact that PMC generates huge amount of data across various projects under various categories and parameters. This raw data if used appropriately can help generate valuable information for PMC as well as entities external to PMC. The corporation receives around 5000 to 5500 RTI requests each year with a monthly average of 450 requests. The number of RTI requests are an indicator that PMC needs to take some concrete measure to make Data available by using a platform that is accessible to various stakeholders like Citizens, NGO's, Private Organizations, Other Govt. organizations, academic institutions etc. This data can be used by these various stakeholders for economic, scientific and developmental purposes. Having such a policy in place will make the city of Pune being identified as one of those cities that is Data Driven and Data Sufficient.

While this document is intended at putting in place a formal Data governance mechanism at PMC it is nevertheless noteworthy to mention that PMC has already made its mark in the field of Open Data by hosting its Open Data Portal which is live as on date. This portal portrays a healthy picture as it has a good 50+ data sets currently uploaded. PMC has also appointed a dedicated City Data Officer to ensure that PMC's vision of being Data Driven and Data Sufficient is led by an able minded professional.

To support its open data initiative PMC also has a basic and high level Open Data Policy based on NDSAP 2012. This document is an attempt to further enhance the existing Data Initiative of PMC by having in place a robust, complete and inclusive Open Data Policy.

3. Pune Open Data Policy and Scope

Pune, with the vision of "Freedom of Information" is committed towards expanding the data that it makes available to the public. It is committed to improve the methods and tools, so that the data is available at a higher frequency. Sharing the information empowers more transparency in governance; conveys progressively productive, efficient open administrations & citizen collaboration; and supports more noteworthy open and business use and re-utilization of government data

This Policy document covers principle considerations concerning the use of data (access & utilization of datasets appropriately enabling the privacy management); the principles governing the Open Data Program for Pune; thereby defining the expectations for department participation and governance of this Open Data Program. This policy covers the data collection, creation and maintenance at City level.

3.1. Purpose

The Purpose of this open Data Policy is to set guidelines for incorporating an open data framework and implementing the same into existing and future systems and procedures. This policy will also aid in determining which data sets should be made public, how to make data sets public and how to maintain and modify the existing published data sets.

With the help of published data sets, the city will enable the public to:

- Assist in distinguishing and proposing the viable and proficient answers for government
- Promote imaginative methodologies/activities for social advancement
- Create economic/financial advancement across the city

3.2. Open Data Policy Goals

The major goals of this Open Data Policy are:

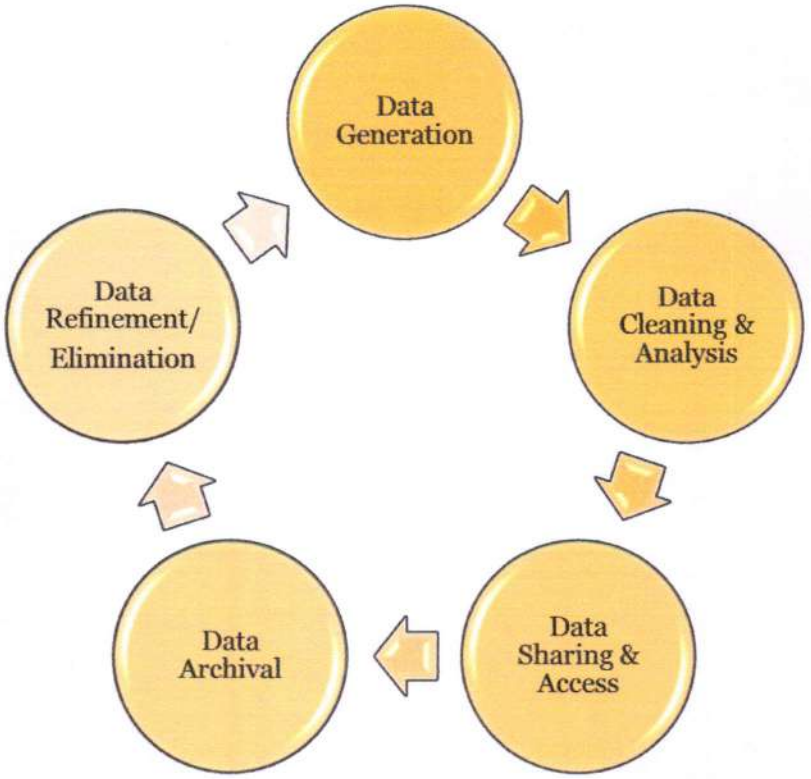
- Data Availability and Accessibility to citizens within defined Legal and Legislative framework
- Improve public understanding of City Operations and other Information concerning their communities
- Spreading a word amongst the citizens to generate economic opportunity for individuals and companies that benefit from the information created by Open Data
- Empower City employees to be progressively effective, better coordinated internally, and recognize chances to better serve the public
- Encourage the development of innovative technology solutions that improve quality of life.

4. Data Lifecycle Management

Implementing a data policy calls for a process that takes care of the Data Lifecycle Management. Having such a process in place will impart a sense of uniformity in the way in which data is

generated/created to the point that it is destroyed/refined/reused. This process will be applicable to all types of data at all levels, all categories and classifications.

The below process diagram showcases the stages in which organizational data should be managed:



- **Data Generation:** PMC's data sets for the open data portal will be generated predominantly from the data received from respective departments of PMC. This data portal is extended to other departments of the city. The raw data may be in the form of excel sheets, word documents, .ppt format, images etc. It will be the CDO's responsibility to develop a template or a form to standardize the data collection/generation activity. Raw data can also be collected by way of interviews, focus group discussions and questionnaires.
- **Data Cleaning and Analysis:** Raw data thus received will have to be processed and analyzed in context to convert it into information which can be effectively used for decision making. Data cleaning will include detecting and removing incorrect records from the received raw data. It also means identifying inaccurate, incomplete and irrelevant parts of the data and replacing/modifying it with meaningful data. Data analysis would mean bifurcation of the data sets into categories and classifications as mentioned in section 4.2.5 below.

- **Data Sharing and Access:** After determining the appropriateness and authenticity of data in the "Cleaning and Analysis" process above the data can be publicly shared by using PMC's digital platforms like Open Data Portal. While sharing these data sets, applicable filtering should be applied to data sets. Also, Security and privacy aspects should be considered while sharing data.
- **Data Archival:** Data, like any applications and products, has its own "end of life" at the arrival of which such data should be archived in a non-production environment and from where it can be again taken back to the production environment if required. Data archival rules can be set on the platform where those data sets which are not being used for a specific period, say 6 months, will be auto-archived in a non-production environment.
- **Data refinement/elimination:** Some data sets by virtue of their nature are highly dynamic in nature and its facts and figures may change at set intervals or at random. Such data sets need to be updated/refined with the latest facts and figures as and when required or at predefined intervals. On the contrary some data sets may become obsolete due their non-relevance to current time, situations and environment. Such data sets, if rendered non-useful, can be eliminated from the data store. The rules of data update, timelines, frequency and rules of elimination to be finalized by the CDO in consultation with department heads.
- **Data Approval Framework:** At every stage of data generation, the concerned stakeholders shall approve and authorize the data usage.

4.1. Data Standards

It becomes imperative to have in place metadata so that available data can be effectively used. This helps achieve uniformity in the way different data features are inferred. The responsibility of adherence to data standards should lie on the shoulders of the City Data Officer (CDO). Important features of metadata are:

- **Standardized name of the data element:** A commonly known and accepted name by which the data element is called across the organization. In case there are different nomenclatures for the same element, the same should be formally documented.
- **Standardized definition of the data element:** Synonymous to the standard name the data element should also have a known and formally accepted definition that will be used to define the element across the organization.
- **Rules for Data Quality:** The data should meet the norms of quality which may include but are not limited to, Accuracy, precision or resolution, completeness, format, consistency, range, pattern etc.

4.2. Data Categorization

Data can be categorized in the below two categories;

- **Personal Data:** Personal data is that data which is specific to a particular living individual. Even incomplete portions of information when combined gives or leads to the identification of an individual is also termed as personal data.
  It hence becomes enormously commanding to 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 Data Sets. 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 with any living individual. Anonymous data is also Non-personal data when all personal indicators and identifiers are eliminated for that particular data element.

4.3. Data Classification

Personal and Non-Personal data can be further classified into the following:

Classification	Data Category	Definition
Level 1	Public	Data available for public consumption and use.
Level 2	Internal Use	Information which can only be disclosed to municipal corporation employees for managing operations or delivery of public services on a day to day basis
Level 3	Sensitive	Data regulated by any city/ State/Central law or regulation like privacy law etc.
Level 4	Protected	Data which needs to be protected e.g. Identity of citizens and disclosure /notification needs to be issued by the

Classification	Data Category	Definition
		municipal corporation in case of any breach or loss of data.
Level 5	Restricted	Data which could lead to a threat to life or loss of public assets or critical infrastructure and are accessible only through a prescribed process of registration and authorization by respective departments/ organizations.

4.4. Data Security and Privacy

In the midst of digital extortion and dangers; Data security and Privacy winds up a standout amongst the most significant factors to be considered while taking a shot at any information strategy or rule. What's more, more so when the information is identified with community organizations and its different activities and undertakings? It is the responsivity of City Data personnel/s to build up a data secured ecosystem. The organization should set down clear standards on the Do's and Don'ts for the entrance and utilization of Data and Information. Accentuation/emphasis ought to be laid on for what reason the information can be utilized and for what reason it can't be. Legitimate activities, so appropriate, must be pre-defined for clients who don't fit in with these guidelines of information use.

4.5. Updating and Maintenance of Data Sets

While uploading of Data Sets forms a major activity of implementing the open data policy, it is equally important to update and maintain these data sets. The following activities should be carried out to ensure the data sets remain relevant and valid.

- **Data Set Updation:** The CDO will ensure that the Data sets are updated whenever it is observed that changes/amendments have occurred in the source of their data i.e. their primary data. It should also be ensured that errors and inconsistencies are removed from the primary data and that the data set included all the metadata along with referential information wherever necessary.

- **Support:** CDO will provide required technical and non-technical support over the queries/inputs/ suggestion received from users through email, portal or social media platforms like Facebook, Twitter and also to internal queries received from PMC's internal departments as well as other departments in the city.
- **Ownership:** All datasets/ feeds remain the property of the organization but are bound by the open data guidelines. The CDO will endorse Government Open Data License to ensure that published data is not misused or misinterpreted by its users.
- **Terms of Use:** Terms of Use are published to restrict the misuse of data and indemnify the city administration in case of any misuse by the end user. The below pointers will be of prime importance while considering "Terms of Use".
  a) All data published on PMC Open Data Portal will remain the property of Pune City. No entity, individual or organization, shall claim ownership of the data sets.
  b) Wherever dataset/s are being used for publication the data's origin must be acknowledged and mentioned.
  c) The datasets published on the Open Data Portal are provided for free of cost, unless otherwise mentioned, on an "as is" and "as available" basis and without any warranty on their accuracy, consistency and suitability for any specific purpose.
  d) PMC or any other departments will not be held liable for any undesired or misfortunate events occurring as a result of the use of such data for purposes that are out of the knowledge of data source. Also, PMC and other government entities will not be liable to pay for any man and material damages caused due to the use of these data sets.
  e) The data sets published shall not be, in any case, be sold or traded by any individual or organization to another individual or organization. Any individual or organization violating this rule will be liable to legal procedures that fall under the jurisdiction of law of land.

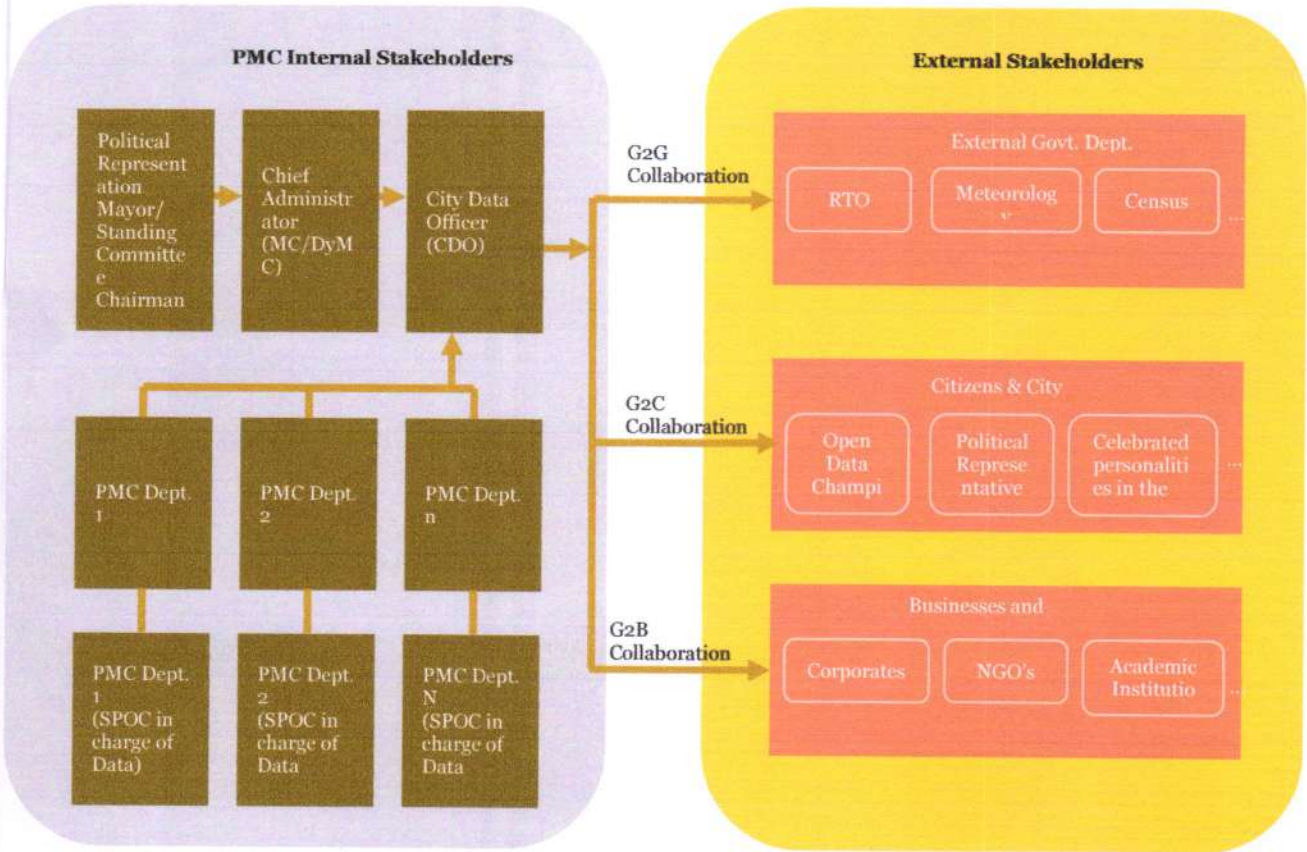
5. Data Set Provisioning and Identifying Data Sets

Pune City has envisioned to have an Open Data Policy in place should critically identify all the important data sets that will be published on their Open Data Portal. These data sets are perceived to have a direct or indirect impact on the citizens and other stakeholders in terms of civic services and organizational administration. Hence it becomes necessarily important to exhaustively identify the data sets that would have varying level of impact on citizens and other stakeholders of the city. While deciding on selecting the data sets the below factors can be considered for.

Need	Influence	Data Market Potential	Legal Implications	Evading Conflicts/Opposition
• Identify the need and demand for the Data Set under consideration. • Engage the important stakeholders demanding this Data Set. • Seggregate the so identified data as Restricted, Non-Restricted, Sensitive, Personal etc.	• Identify the influence that any data set would have post influencing. • Data sets with negative influence should be avoided from being published. • Data sets that could have positive influence and impact on socio-economic factors like jobs and employment, education and litercay should be published.	• Data sets possess an inherent potential to generate economic value • Such data sets that could boost startups and entrepreneurship, innovation and change should be priotrized for publishing. • Data sets with potential to boost the city's economy in any and all ways should be treated as High Value Data sets.	• Data sets that have a legal binding to be published, for instance under the ambit of RTI, should be identified and made public. • PMC receives around 5000 RTI requests annually. PMC should proactively identify critical Data Sets so that RTI requestes and Queries are reduced and hence legal compliance is adhered.	• Publishing sensitive data sets can attarct undue political attention. • Some Data sets may have the potential to expose unattended civic issues and problems and hence bring to fore the people/dept. responsible for these issues. • Utmost care shuold be taken to ensure that publishing such data sets does not hamper the continuity of the open data inititive.

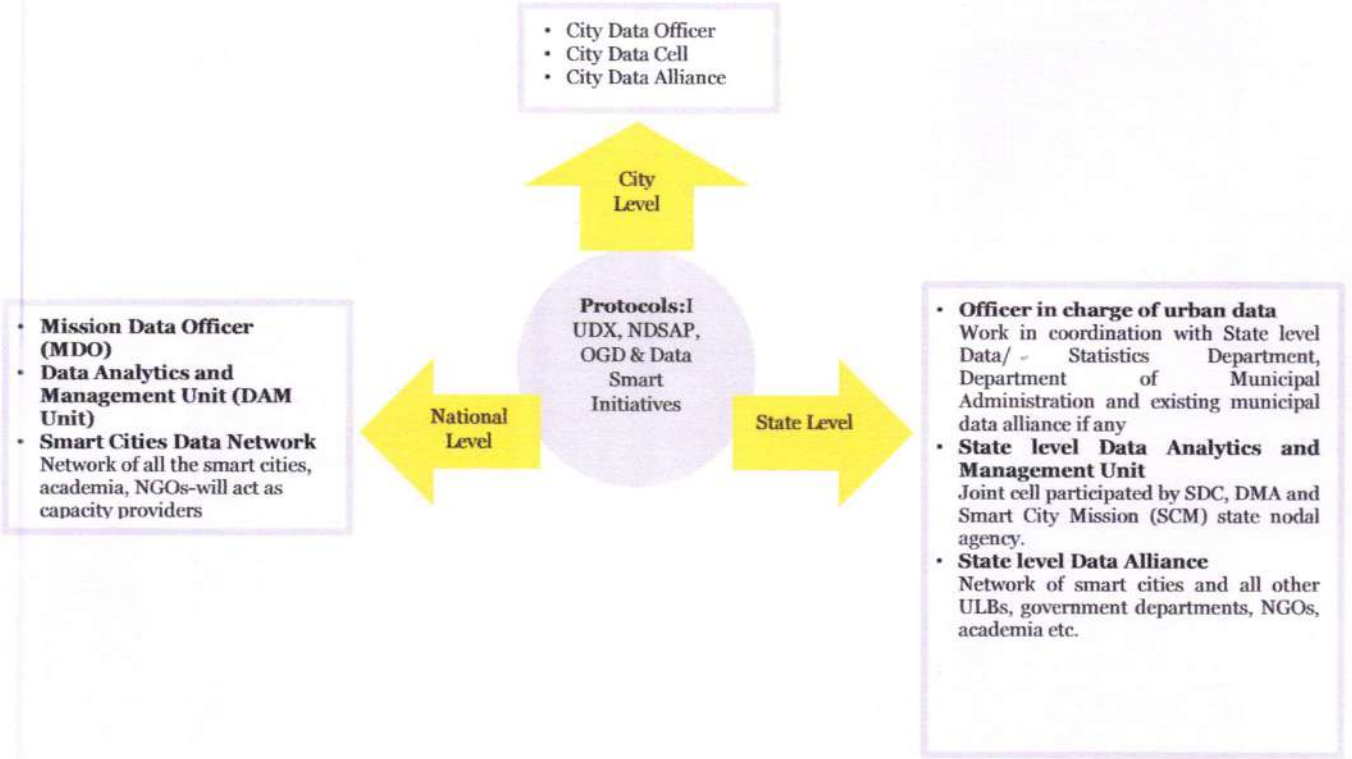
6. Stakeholders and Collaboration

In order to successfully implement the Open Data Initiative it will have to identify the various stakeholders within PMC and external to it. After having identified them it is equally important to collaborate effectively with these stakeholders. In initiatives like these it is of utmost importance to successfully establish Government to Government (G2G), Government to Citizen (G2C) and Government to Business (G2B) collaborations.



7. Data Management and Team Structure

Though Open Data Policies may be written for a specific city or ULB under consideration, yet there is always going to be the need of close collaboration and integration with State Level and National Level entities. The below diagram depicts the various entities involved at various levels that adhere to set protocols and guidelines as prescribed by GoI.



8. Roles and Responsibilities

8.1. City Data Officer (CDO)

The CDO will act as custodian and driver of City Data Policy (CDP) and a flag bearer of open government initiative. CDO's major responsibility is to put data to its right use i.e. for generating insights, using data for effective service delivery or infrastructure delivery, improving civic operations by making real time decision making etc. City Data Officer will report directly to city leadership and act as single point of contact to all internal and external stakeholders in the city. CDO will be supported by a pool of Data Coordinators. The CDO will work with all the departments including those which are outside the jurisdiction of the municipal corporation. For example, Traffic, Police, power utilities, telecom networks, gas distribution, etc. Following are the roles and responsibilities of CDO:

- The CDOs will create a City Data Policy (CDP) for Pune Smart City which will be reviewed every month to keep it contextual to the need of the times.
- Coordinate with MDO to align with mission data strategy and priorities with respect to Open government initiatives and policies. City Data Officer along with a team of data champions/ coordinators must assess the data requirements of various stakeholders in the smart city ecosystem.
- Coordinate with MDO to align with mission data strategy and priorities with respect to Open government initiatives and policies.
- Coordinate with officers of various other government departments/agencies within the city for the effective implementation of City Data Policy.
- Organize regular meetings of Smart City Data Alliance (SCDA).
- City Data officer will engage with the external stakeholders to understand the data requirements of the city.
- City Data Officer must identify cross-cutting data sets and publish it on a platform based on data needs of various stakeholders in a routine manner.
- City Data Officer must design a program to solve its problems through a structured challenge process like hackathon, events etc.
- City data officer 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
- Publish Data Catalogues and Data Sets/Feeds on OGD portal: CDOs will publish data Catalogues and Data Sets/Feeds on OGD Portal and will ensure that such data sets are updated at regular time intervals as needed and create mechanisms for continuous feedback from citizens and stakeholders on type of data sets to be published. The CDO will be responsible for publishing of such data sets/ feeds as mandated as part of Mission Data Strategy.

8.2. Open Data Champions

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 City Data Policy at the department level. DCs needs to act as trainers and lead the team of data coordinators at the department level. DC will be first touch point of CDO and must undertake continuous capacity building programs for their staff. Following are the roles and responsibilities of Data Champions:

- Shall identify the data sets/feeds, derived information, intelligence or data challenge with respect to day to day operations of the department.
- Actively publish/ enable to publish data sets/feeds identified as relevant to the resolution of critical use cases for the city. They will work closely with the CDO for active implementation of the City Data Policy.
- 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.
- DCs will undertake activities to engage with their stakeholders and evolve their department's strategy on data in line with the deliberations.

8.3. Open Data Coordinators

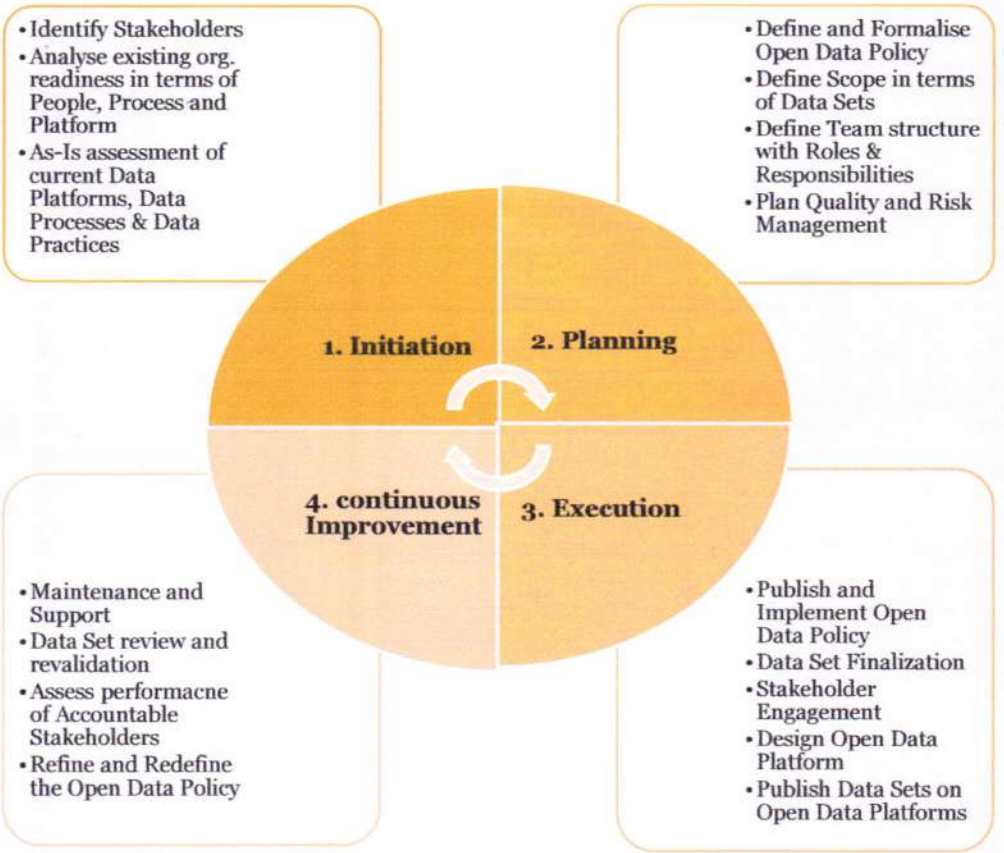
Data Coordinators are designated for each department of PMC as the main point of contact and accountability for open data in their department and are sometimes the individual in charge of individual database, data sets or information systems. Data Coordinators will assist the Data Champions at the department/ PMC Enterprise level as reporting staff. These Data Coordinator will act as a data management experts who will coordinate, gather, collect and sanitize the data for Open Data Portal development. Overall data coordinator is the person who knows how to get access to data. For the successful roll out and implementation of Open Data Policy, set of roles and responsibilities to be defined for Data Coordinators. To achieve the same, following shall be the responsibilities of appointed data coordinators:

- Aggregate the data demand from various channels
- Serving as a key point of accountability for timelines and questions about data sets
- Implementing privacy, metadata and other standards & practices
- Sensitizing the department employees over the importance of data quality etc.
- Support to Data Champion for implementation of city data policy at department level
- Understand the public demand to identify high value data sets so as to make them available in a way that maximizes the number of users and portal usage

- Assist the data champion and city data officer for continuous updation and improvement in open data policy
- Provide the guidance and training to department users regarding open data portal and its usage
- Continuous support and coordination with concerned stakeholders for effective implementation of Data Maturity Assessment Framework
- Engage with other departments to prototype and execute data-related projects
- Maintain data integrity and security and ensure the proper sorting and organizing of data sets
- Should prioritize and classify the data as per the policy standard. e.g. Data to be categorized as Personal or Non-Personal Data, further it should be classified in the defined category like Secured, Protected, Restricted data
- Collaborate with other public sector organizations/departments to define best practices and lessons learned for similar projects
- Data updation and maintenance on periodic basis based on the timelines/granularity defined by PMC
- Assist with other key open data initiatives led by the PMC, MoUHA or any other government bodies in coordination with CDO and Data Champion
- Inventorying department data sets
- Managing accuracy, quality and completeness of the data
- Assisting departments and Data Champion with analysis of city data sets
- Determining what datasets are appropriate for public disclosure
- Managing access to and use of the data, including documentation in coordination with CDO and Data Champion

9. Open Data Policy – Implementation Plan

Once Open Data Policy is approved and formalized, the next step for PMC would be to implement the Policy while adhering to the norms and guidelines as stated in this policy document. Implementing an Open Data policy may not be a simple process as it involves the participation of various stakeholders like Civic Administrators, Political Champions, external departments etc., as mentioned in section 5.1 above. Apart from having a Data Policy in place it also becomes equally important to have an implementation plan and roadmap in place to successfully implement the open data initiative. The below diagram showcases a referential implementation plan that complies while implementing its Open Data Policy.



10. Standard Operating Procedures

Standard Operating Procedures (SOPs) are succinct formal documents designed to achieve consistency in specified functions by specifying standard practice in performing those functions. These SoPs are designed with a view to enhance and standardize its open data operations:

1. SoP for data collection
2. SoP for electronic data collection
3. SoP for data processing and cleaning
4. SoP for data for quality assessment of data sets
5. SoP for data publishing as per Open Data Norms
6. SoP for engaging stakeholders to assess the data needs
7. SoP for data collection, processing and analysis for on field Survey

In order to have responsibilities defined, RACI Matrix defined. It is a powerful tool to assist in the identification of roles and assigning of cross-functional responsibilities.

RACI Definitions:

Responsibility = person or role responsible for ensuring that the activity is completed

Accountable = person or role responsible for actually doing or completing the activity

Consulted = person or role whose subject matter expertise is required in order to complete the activity

Informed = person or role that needs to be kept informed of the status of activity completion

10.1. SoP for Data Collection

The purpose of this SoP is to define the Data collection processes in the collection and management of open data held or on behalf of Pune City and the responsibilities of individuals involved. It also aims to describe good practice in data collection and management techniques. This SOP should be used when any form of data is collected, accessed, transferred or stored.

Stakeholders / Open Data Activities	Municipal Commissioner	Chief Data Officer	City Data Officer	Data Champion	Data Coordinators	Other Representatives (Alliance/ Dept. SPOC etc.)
Identification of data sets	I	I	A	C	R	I/C
Defining standard format or data set format for collection/gathering of data	I	I	A	C	R	I/C
Data collection	I	I	R	C	A	C
Data sanitization	I	I	A	C	R	I/C
Data categorization and classification	I	I	A	R/C	R	I/C
Data Standardization	I	I	A	R/C	R	I/C
Data Approval	I	A	R	C	I	I
Publishing the data sets	I/C	C	A	I	R	I

10.2. SoP for Electronic Data Collection

The purpose of this SoP is to define the Electronic Data collection processes in the collection and management of open data held or on behalf of Pune City and the responsibilities of individuals involved. It also aims to describe good practice in data collection and management techniques electronically. This SOP should be used when any form of data is collected, accessed, transferred or stored.

Stakeholders / Open Data Activities	Municipal Commissioner	Chief Data Officer	City Data Officer	Data Champion	Data Coordinators	Other Representatives (Alliance/ Dept. SPOC etc.)
Identification of data sets	I	I	A	C	R	I/C
Defining standard format or data set format for collection/gathering of data	I	I	A	C	R	I/C
Data collection	I	I	R	C	A	C
Data sanitization	I	I	A	C	R	I/C
Data categorization and classification	I	I	A	R/C	R	I/C
Data Standardization	I	I	A	R/C	R	I/C
Data Approval	I	A	R	C	I	I
Publishing the data sets	I/C	C	A	I	R	I

10.3. SoP for Data Processing and cleaning

The purpose of this SoP is to define the Data processing and cleaning while doing the collection and management of open data held or on behalf of Pune City and the responsibilities of individuals involved. It also aims to describe good practice in processing of data and cleaning techniques. This SOP should be used when any form of data is collected, accessed, transferred or stored.

Stakeholders / Open Data Activities	Municipal Commissioner	Chief Data Officer	City Data Officer	Data Champion	Data Coordinators	Other Representatives (Alliance/ Dept. SPOC etc.)
Data identification and Collection	I	I	R	R/C	A	C
Defining standard format or data set format for collection/gathering of data	I	I	A	C	R	I
Data Sanitization	I	I	A	C	R	I/C
Data Classification and standardization	I	I	A	C	R	I/C
Analysis and Accuracy Validation	I	I	A	C	R	I/C
Periodic gathering and collection of datasets as per defined frequency and granularity	I	I	R	R/C	A	I/C
Security Check involving removal of personalized data	I	C	A	C	R	I/C

10.4. SoP for Data for Quality Assessment and Data Sets

The purpose of this SoP is to define the processes for the quality assessment and data sets of open data held or on behalf of Pune City and the responsibilities of individuals involved. It also aims to describe good practice in quality assessment and Datasets management. This SOP should be used when any form of data is collected, accessed, transferred or stored.

Stakeholders / Open Data Activities	Municipal Commissioner	Chief Data Officer	City Data Officer	Data Champion	Data Coordinators	Other Representatives (Alliance/ Dept. SPOC etc.)
Data collection as per the standards defined	I	R	C	R/C	A	I/C

Data sanitization	I	I	A	C	R	I/C
Security Check involving removal of personalized data	I	C	A	C	R	I/C
Units/Currency and other terms standardization	I	C	A	C	R	I/C
Periodic gathering and collection of datasets as per defined frequency and granularity	I	I	R	R/C	A	I/C

10.5. SoP for Data Publishing as per Open Data Norms

The purpose of this SoP is to define the data publishing processes as per the Open Data Norms of open data held or on behalf of Pune City and the responsibilities of individuals involved. It also aims to describe good practice in data publishing and management techniques as per Open Data Norms. This SOP should be used when any form of data is collected, accessed, transferred or stored.

Stakeholders / Open Data Activities	Municipal Commissioner	Chief Data Officer	City Data Officer	Data Champion	Data Coordinators	Other Representatives (Alliance/ Dept. SPOC etc.)
Data categorization and classification	I	I	A	C	R	I
Data Standardization	I	I	A	C	R	I
Analyzing the data	I	I	A	C	R	I
Consulting and identifying new data sets	I	C	R	R/C	A	I/C
Defining new problem areas to tackle upcoming problems	I	C	A	C	R	C
Constantly enhancing the quality of data sets	I	C	A	C	R	C
Keeping concern stakeholders updated about the new guidelines and open data norms	I	I	A	C	R	C/I

10.6. SoP for Engaging Stakeholders to Assess the Data Needs

The purpose of this SoP is to define the processes for engaging stakeholders to assess the data needs of open data held or on behalf of Pune City and the responsibilities of individuals involved. It also aims to describe good practice in engaging the stakeholders to assess the data needs. This SOP should be used when any form of data is collected, accessed, transferred or stored.

Stakeholders / Open Data Activities	Municipal Commissioner	Chief Data Officer	City Data Officer	Data Champion	Data Coordinators	Other Representatives (Alliance/ Dept. SPOC etc.)
Organizing workshops, hackathons/events and trainings	I/C	C	A	C	R	I/C
Decision making and consultation with data experts	I	C	A	I	R	I/C
Data ideation with public forum	I/C	C	A	C	I	I/C
Formation of city data alliances	I/C	C	A	R	I/R	I/C
Identifying stakeholders from various age groups and ethnicities and engaging them in city initiatives	I	C	A	R	I/R	I/C
Brainstorming and role modelling sessions with concerned stakeholdes and subordinates	I	C	A	R	I/R	I/C

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

The purpose of this SoP is to define the Data collection, processing and analysis processes for on field survey of open data held or on behalf of Pune City and the responsibilities of individuals involved. It also aims to describe good practice in data collection, processing and analysis techniques for on field survey. This SOP should be used when any form of data is collected, accessed, transferred or stored.

Stakeholders / Open Data Activities	Municipal Commissioner	Chief Data Officer	City Data Officer	Data Champion	Data Coordinators	Other Representatives (Alliance/ Dept. SPOC etc.)
Electronic device usage for identification and collection of Data	I	C	A	C	R	A/C
Quality assessment of data	I	C	A	C	R	I/C
Defining standard format or data set format for collection/gathering of data	I	I	A	C	R	C
Data Sanitization	I	I	A	C	R	I/C
Data Classification and standardization	I	I	A	C	R	I/C
Measuring and reviewing the accuracy of data	I	I	A	C	R	I/C
Identification of new data sets using electronic devices	I	C	A	R/C	R	C
Periodic gathering and collection of	I	I	A	R/C	A	I/C

datasets as per defined frequency and granularity						

10.8. Data Analysis for collected data

The purpose of Data analytics is to derive conclusion, trends, analytical insight about the information contained in the datasets collected through multiple process and procedures by examining them. Data analytics will enable Pune city administration and stakeholders to take raw data and uncover patterns to extract valuable insights from it, which in turn will enable the administration to take necessary actions.

Type of analytics that will be performed over the collected data will be:

Descriptive Analytics: The purpose of this analytics type will be to summarize the findings and understand what is going on.

Diagnostic Analytics: It will be used to determine why something happened in the past.

Predictive Analytics: These analytics is used to predict future outcomes. It focuses and forecasts what are the probabilities of the occurrence of the event.

Prescriptive Analytics: This type of analytics suggest all favorable outcomes according to a specified course of action and also suggest various course of actions to get to a particular outcome

Annexures

1. Open Government Data Platform Policy

Government of India has listed down policy for external users to access the data hosted on it Open Government Data Platform. This policy can be access at https://data.gov.in/policies. A summary of such policy is presented as following, the same needs to be implemented and followed by Pune while implementing the policy document –

Sr. No.	Policy Name	Details
1.	**Hyperlinking Policy**	Open Government Data(OGD) Platform, India do not object the user linking directly to the information that is hosted on the Open Government Data Platform India (http://data.gov.in) and no prior permission is required for the same.
2.	**Links to external websites/portals**	User would find various sections of the ODG Platform website links to other websites/portals. These links have been provided to enhance user's experience and are presented for information purposes only
3.	**Privacy Policy**	User's privacy is important to ODG and Govt. of India. User is allowed to visit the site without revealing Personal Information unless he chooses to provide such information. **ODG Platform collects and stores the following information about user visits for statistical purpose:** • The internet domain of users services provider and IP address from which user access our website. • The type of browser and operating system used to access our site. • Date and time user accessed our site. • The pages/URLs visited • If the visitor reached this website from another website, the address of that referring website. **Personal Information**