

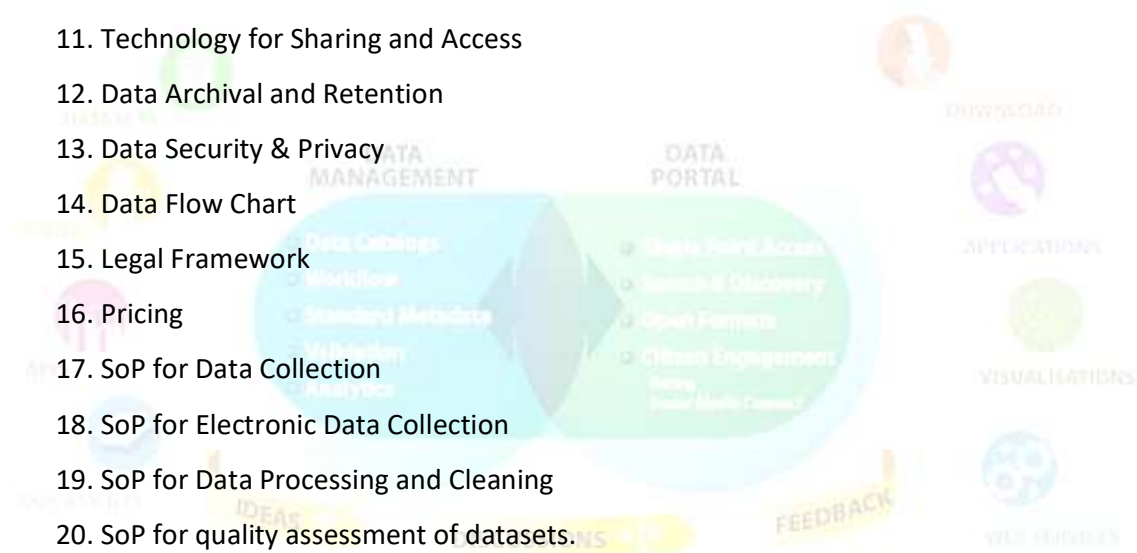


City Data Policy



Table of Content

1. Introduction
2. Objective
3. Definition
4. Scope of the Policy
5. Benefits of the Policy
6. Stake Holders
7. Data Classification
8. Data Categorizations
9. Type of Access
10. Roles & Responsibilities
11. Technology for Sharing and Access
12. Data Archival and Retention
13. Data Security & Privacy
14. Data Flow Chart
15. Legal Framework
16. Pricing
17. SoP for Data Collection
18. SoP for Electronic Data Collection
19. SoP for Data Processing and Cleaning
20. SoP for quality assessment of datasets.
21. SoP for engaging stakeholders to assess the data needs
22. SoP for data collection, processing, and analysis for on field survey



Introduction:

The production of data in urban environments is constantly increasing in the last few years due to the deployment of smart city technologies and a generalised awareness of the value of data as “engines for the new economy”. Nevertheless, the unregulated collection and exploitation of data produced in cities (city data) is fostering the rise of significant political and social concerns. Indeed, data are not only resources but also a source of conflicts in the relationships between government and citizens, technology providers and users clients, organisations capitalising the access to information, and people excluded from the benefits generated from their data.

City Data Policy is the first significant step in the direction to provide conceptual clarity over accessing and sharing protocols over city data. Data Policy must address concerns like data classification, categorization, archival, security, privacy, ownership etc. The absence of a City Data Policy also acts as a barrier towards setting up a data economy at the city level. City Data Policy provides clarity around ownership of data, legal framework, terms of use, etc.



Objective:

The objective of this policy is to facilitate Agra Smart City Limited owned sharable data and information in both human readable and machine readable form through a network all over the city in a proactive and periodically update manner, with in the framework of the various related policies, Act and rules of Government of India and Government of Uttar Pradesh, thereby permitting wider accessibility and use of public data and information.



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.

Data Set means statistical or factual information (1) contained in structured data sets; and (2) regularly created or maintained by or on behalf of and owned by a department which supports or contributes to the delivery of its services, programs, and functions. Such term shall not include image files or other non-structured data sets. Nothing in this policy shall be deemed to prohibit a department from voluntarily sharing information not otherwise covered under this policy.

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.

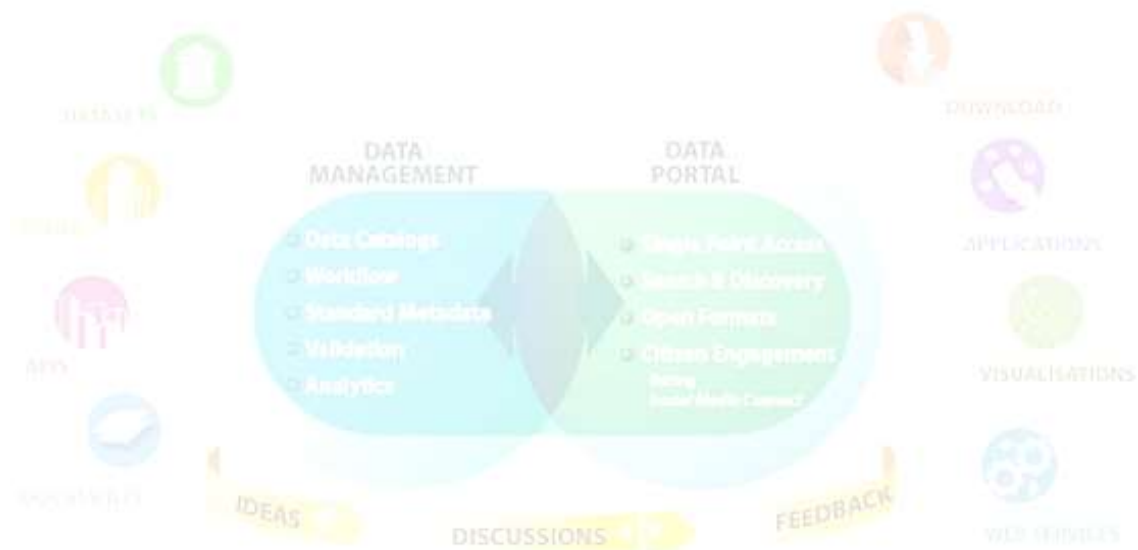
Restricted Data: Data which are accessible only through a prescribed process of registrations and authorization by respective departments/organisation since it could lead to a threat to life or loss of public assets or critical infrastructure.

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 Data Portal means an online repository for open data that is assessable through the public Internet to facilitate access to and re-use of public sector information.

Open Data means public information that should be freely available for use and re-use that does not need to be kept private due to federal law, state statute, city ordinance, or other policy.



Scope of the Policy:

This policy will apply to all data and information created, generated, collected and archived using public funds provided by various department of Central Government and State Government.

This Policy document covers principle considerations concerning the use of data (access and utilisation of datasets including appropriate privacy management), and the principles governing the data sharing program for Agra; thereby defining the expectations for departmental participation and governance of the data program.

The policy is intended as a resource for city administrators such as Municipal Commissioner, Smart City CEO, and other officials such as City Data Officer, heads of various government departments, Data Coordinators, Data Champions and external agencies - (parastatal, civic, private) interested in engaging with the data initiatives of the City.

This Policy will apply to all data and information created, generated, collected and archived by Agra Municipal Corporation/Agra Smart City Limited. This policy applies to any person/user, organisation, administrators, contractors, etc. who intends to access information or assets through any data portal of Agra Municipal Corporation/Agra Smart City Limited. Specifically, the Data Policy applies to the following information assets of Agra Municipal Corporation/Agra Smart City Limited:

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

Benefits of the Policy:

Maximising Use: Ready access to government owned data will enable more extensive use of a valuable public resource for the benefit of the community.

Avoiding Duplication: By sharing data the need for separate bodies to collect the same data will be avoided resulting in the significant cost saving in the data collection.

Maximised Integration: By adopting common standards for the collection and transfer of data, integration of individual datasets may be feasible.

Ownership Information: The identification of owners for the principal data sets provide information to identify those responsible for implementation of prioritized data collection programs and development of data standards.

Better decision-making: Data and information facilitates making important decision without incurring repetitive cost. Ready access to existing valuable data is essential for making decision making task such as protecting the environment, development planning, managing assets, improving living conditions, security and controlling disasters.

Equity of access: This policy ensure better access to all bonafied users.



Stakeholders:

The different classes of stakeholders operating in Agra as Users of the solution. They include:

- (a) Local governments (political and administrative authorities at the city level)
- (b) Public agencies (administrative and operational unit for the provision of public services at local, regional and national level)
- (c) Business sector (industry, commerce, private services)
- (d) Start-ups and Incubators: Representatives from start-ups and incubators in the city.
- (d) Non-profit organisations (charities, foundations)
- (e) Civil society groups (citizens and informal groups)
- (f) Knowledge cluster (universities, research centres, R&D departments of public and private organisations, knowledge professionals).
- (g) Professional Representatives: Representatives from various professional services like Doctors, CA, and Engineers etc.



Roles & Responsibilities

City Data Officer (CDO)

The CDO will be the officer responsible for implementation of the Data Smart Cities Strategy at the city level. CDO will work with city leadership to assess and tap the potential of data and set up data culture across the organisation and outside the organisation. CDO will report directly to City Leadership and act as single point of contact to all internal and external stakeholders in the city. City leadership also needs to deploy dedicated skilled resources to drive the data initiative through CDO.

The key responsibilities of CDO are as follows:

- a. Ensure that the CDP evolves as per the needs of various stakeholders of the city and relevant upgrades to policy are carried out time to time accordingly.
- b. Coordinate with MDO (Mission Data Officer) to align with mission data strategy and priorities with respect to open government initiatives and policies.
- c. Organise regular meetings of the City Data Alliance (CDA).
- d. Coordinate with officers of various other government departments/agencies within the city for the effective implementation of CDP.
- e. CDO along with team of Data Champions/ Coordinators must assess the data requirements of various stakeholders in smart city ecosystem. External stakeholders may also need to be engaged to understand the data needs. CDO must engage various internal stakeholder at operational, tactical and strategic level to assess the data need to make decisions. Data needs and frequency of consumption needs to be outlined for internal stakeholders.
- f. Publish Data Catalogues and Datasets/Feeds on Open Government Data Portal and ensure that such datasets are updated at regular intervals as needed and create mechanisms for continuous feedback from citizens and stakeholders on type of datasets to be published.
- g. Assess all the operational IT Projects for identifying public datasets/feeds. Data Champions and Data Coordinators in respective departments must prepare integration plan with respective IT vendor/ integrator to ensure compliance as per CDP.

- h. 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. CDO must work with concerned System Integrator/vendor to ensure compliance of smart solutions with CDP.
- i. Assess all periodic and recurring MIS needs to identify the datasets/feeds which could be shared with other departments through data exchange. CDO will also assess third party funded reports related to city operations for e.g. City Mobility plan, Health Plan etc.

Data Champions (DCs)

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

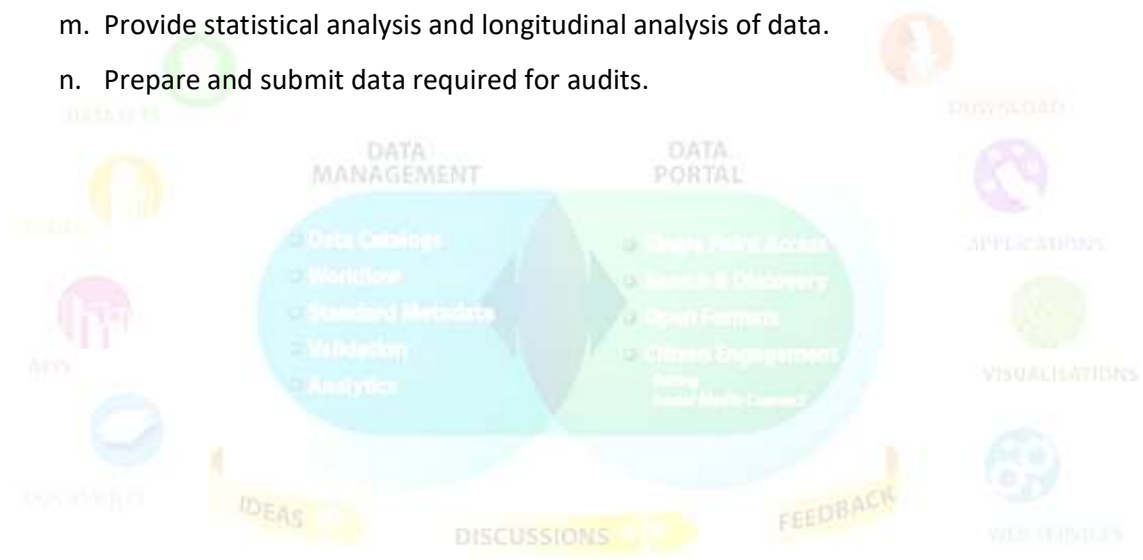
- a. Shall identify the datasets/feeds, derived information, intelligence or data challenge with respect to day to day operations of the department.
- b. Actively publish/enable publishing of datasets/feeds identified as relevant to the resolution of critical use cases for the city. They will work closely with the CDO for active implementation of the CDP.
- c. 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.
- d. Undertake activities to engage with their stakeholders and evolve their department's strategy on data in line with the deliberations.

Data Coordinators

Data Coordinators will assist Data Champions at the department/government agency level as reporting staff. Their responsibilities are as follows:

- 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 Agra Municipal Corporation/Agra Smart City Limited or CDO via Internet.
- g. Update Agra Municipal Corporation/Agra Smart City Limited website 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 Agra Municipal Corporation/Agra Smart City Limited 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.



Data Classification:

Different types of data sets generated both geospatial and non-spatial for by different departments are to be classified as shareable data and non sharable data. The types of data produced by a stastical system consist of derived statistic like accounts statistics, indicators like price index, data bases from surveys. The geospatial data however, consist of primarily of satellite data, maps etc. In such a system, it becomes important to maintain standards in the respect of metadata, data layout and data access policy.

The connection between Users (city stakeholders) and Object (city data) had been explored in terms of relationships between city components and classes of local stakeholders related to their role in the city data ecosystem as:

- (a) Data collectors, gathering data in the context
- (b) Data processors, analysing and structuring data
- (c) Data providers, managing the distribution of data
- (d) Data consumers, using data
- (e) Data subjects, actively or passively providing data on their activities

Data Categorization:

Data is classified into three categories:

- A. Restricted Data:** Data should be classified as Restricted when the unauthorized disclosure, alteration or destruction of that data could cause a significant level of risk to the Institution or Urban Local Body. The highest level of security controls should be applied to Restricted data.
- B. Private Data:** Data should be classified as Private when the unauthorized disclosure, alteration or destruction of that data could result in a moderate level of risk to the Urban Local Body. By default, all Institutional Data that is not explicitly classified as Restricted or Public data should be treated as Private data. A reasonable level of security controls should be applied to Private data.

C. Public Data: Data should be classified as Public when the unauthorized disclosure, alteration or destruction of that data would result in little or no risk to the Urban Local Body. Examples of Public data include press releases, departments information and tenders publications etc. While little or no controls are required to protect the confidentiality of Public data, some level of control is required to prevent unauthorized modification or destruction of Public data.

Types of Access:

Open Access: Access to data generated from public funding should be easy, timely, user-friendly and web-based without any process of registration/authorization.

Registered Access: Datasets which are accessible only through a prescribed process of registration/ authorization by respective department/ organizations will be available to the recognized institutions / organizations/ public users, through defined procedures.

Restricted Access: Data declared as restricted, Central Government/ State Government policies, will be accessible only through and under authorization.

Technology for Sharing and Access:

A state-of-the-art data warehouse and data archive with online analytical processing (OLAP) capabilities, which include providing, a multi-dimensional and subject oriented view of database need to be created. The main features of the data warehouse need to include:

- a) User friendly interface
- b) Dynamic/ pull down menus
- c) Search base reports
- d) Secured web access
- e) Bulletin Board
- f) Complete Metadata
- g) Parametric and Dynamic report in exportable format.

Roles and Responsibilities for Administrating the Policy

The publication of data sets will be done in close cooperation between departments and the IT Section, according to the following guidelines:

Department Responsibilities	IT Section/Department Responsibilities
Identify data that should be designated private on account of valid privacy, security, privilege, statutory, administrative cost concerns, or other privacy reasons and, if possible, specify how any private data can be aggregated, generalized or otherwise de-identified so it can be made public.	Provide and administer a reliable data portal adequately provisioned to host department/city data.
Maintain accuracy and quality of supplied data sets.	Publish Department data through the department/City web site accessible to external search capabilities.
Support (i.e. fund) development necessary to extract, transfer and load data.	Develop necessary technical processes, automated if possible, to extract, transfer and load department data.
Provide a comprehensive inventory of information possessed and/or managed by the Department, classifying data sets as currently "public" or "not yet public" or "protected" (give reasoning).	Provide an Open Data catalogue capability.
Assist in the resolution of discrepancies or inconsistent reporting results.	Maintain integrity of supplied data sets.
Provide subject matter expertise necessary to interpret data set content for public.	Make provision for support in the use of the City data portal by the public.
Describe the data fields contained in given data set (i.e. metadata).	Establish minimal metadata requirements needed for the public to

	understand what the data are, and how and when they were collected.
Notify IT Section/Department in advance when an application or database will be unavailable, changed or restated.	Implement functionality to solicit feedback from the public and to encourage public discussion on open data policies and public data set availability on the City data portal.



Data Archival and Retention

Archiving is defined as secured storage of data/ documents, such that the same is rendered inaccessible by authorised users in the ordinary course of business, but which can be retrieved by an administrator designated by the HoD for the document in question. Based on certain aspects such as compliance with statutory and regulatory requirements, responses to inspections from regulator and availability of documents for any statutory assessments.

Advantages of data archiving are:

Increased capacity — Archiving digital data ensures backup and recovery runs faster.

Easier backup — Data archiving techniques can also ensure simpler backup processes because you don't waste time backing up inactive data.

Improved ability to meet compliance requirements — Regardless of your industry or vertical, data archiving requirements and best practices can ensure your organization stays in compliance with applicable regulations and the law.

Enhanced productivity — Spend less time maintaining and managing software and infrastructure for on-site backup storage.

Higher growth — A scalable, cost-effective cloud data archiving solution allows for a pay-as-you-go growth mode without as much waste, even in industries that generate high amounts of data.

More refined management of locations — Using a virtual data archiving system allows for savings on investments into office intranets and other costly infrastructure.

Mode of Archival:

Electronic: Records to be archived in electronic form. These records may be stored on assigned servers or on tapes as per rules and guidelines issued by the Head of the department.

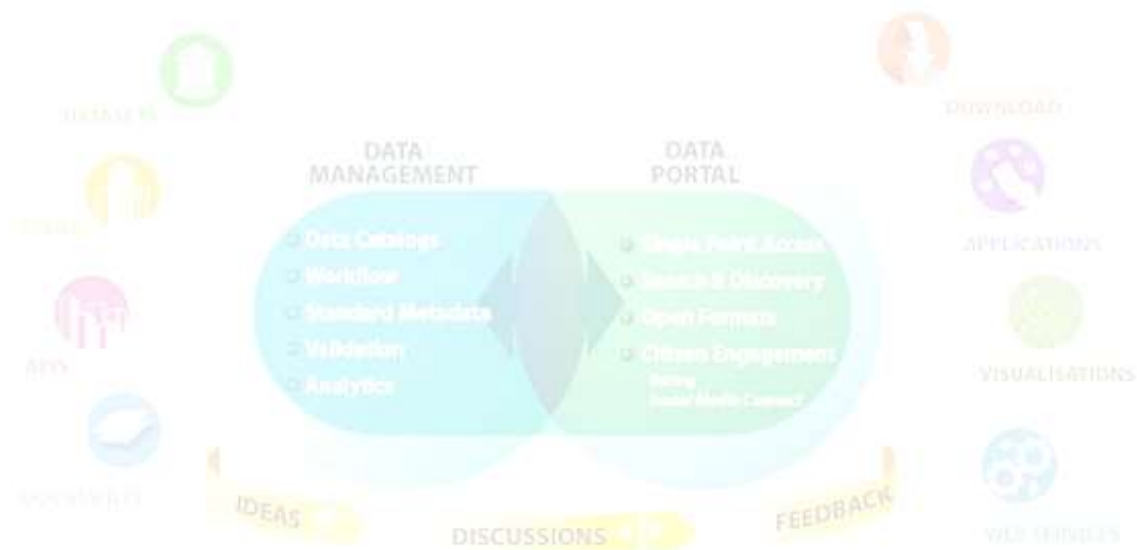
Physical: Record to be archived in physical form. The records may be archived in the concern officer or record room.

Original Form: Records to be retained in the original form in which they were created or used i.e. either electronic or physical.

Categories for archival of data of Agra Municipal Corporation:

1. Category-I (e-Files/records to preserved permanently which are of historical importance) – up to 5 it will be kept in the Department’s server and thereafter transferred 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) – till 5 years on the Department’s server. In exceptional cases, if the record is required to be retained beyond 10 years it will be upgraded to Category-I.

Data will be kept in local storage for the single Financial year after completions of financial year it will transfer to the data storage devices/ archived.



Data Security and Privacy:

Data security refers to the process of protecting data from unauthorized access and data corruption throughout its lifecycle. Data security includes data encryption, hashing, tokenization, and key management practices that protect data across all applications and platforms.

Each set of data has its own security and privacy consideration, as well as commercial, monetary or subscription aspects which must be observed.

The fundamental principles (tenets) of information security are confidentiality, integrity, and availability. Every element of an information security program (and every security control put in place by an entity) should be designed to achieve one or more of these principles.

Confidentiality measures are designed to protect against unauthorized disclosure of information. The objective of the confidentiality principle is to ensure that private information remains private and that it can only be viewed or accessed by individuals who need that information in order to complete their job duties.

Integrity involves protection from unauthorized modifications (e.g., add, delete, or change) of data. The principle of integrity is designed to ensure that data can be trusted to be accurate and that it has not been inappropriately modified.

Availability is protecting the functionality of support systems and ensuring data is fully available at the point in time (or period requirements) when it is needed by its users. The objective of availability is to ensure that data is available to be used when it is needed to make decisions.

Effectively executing all three tenets of the Security Triad creates an ideal outcome from an information security perspective.

Sensitive data: that will be used in the course of business/services operations. Because the data is sensitive, that data should only be able to be seen by the people in the organization (Agra Municipal Corporation/ Agra Smart City Limited) that need to see it in order to do their jobs. It should be protected from access by unauthorized individuals.

When the individual that needs that piece of data to perform a job duty is ready to utilize it, it must be readily accessible (i.e. online) in a timely and reliable manner so the job task can be completed on time and the Urban Local Body (Agra Municipal Corporation) can continue its processing. This describes the principle of availability.

The data will be used in calculations that affect Urban Local Body (Agra Municipal Corporation) decisions and finance that will be made by the organization. Therefore, the accuracy of the data is critical to ensure the proper calculations and results upon which decisions will be made. The assurance that the data has not been improperly tampered with and therefore can be trusted when making the calculations and resulting decisions is the principle of integrity.

Security Areas which will ensure whenever data will be published are:

1. Physical Security
2. System Security
3. Application Security
4. Network Security
5. Audit Trails & Logs
6. Data Protection
7. Session Management
8. Data Ware house Security
9. Application Deployment Security
10. Compliance of Security Standards

Data Privacy:

Hyperlinking Policy: We do not object to you linking directly to the information that is hosted on the Municipal Corporation/ Smart City website and no prior permission is required for the same. However, we would like you to inform us (using the Contact Us option) about any links provided to our site. We do not permit our pages to be loaded into frames on your site. The pages must load into a newly opened browser window of the user. Please visit Link to Us page for more details. Links to external websites/portals You would find various sections of this website links to other websites/portals. These links have been provided to enhance your user

experience and are presented for information purposes only. Agra Municipal Corporation/ Agra Smart City Limited is not responsible for the contents, usability, accessibility, reliability including cyber security related issues, if any, of the linked websites and does not necessarily endorse the views expressed in them. We cannot guarantee that these links will work all the time and we have no control over availability of linked pages.

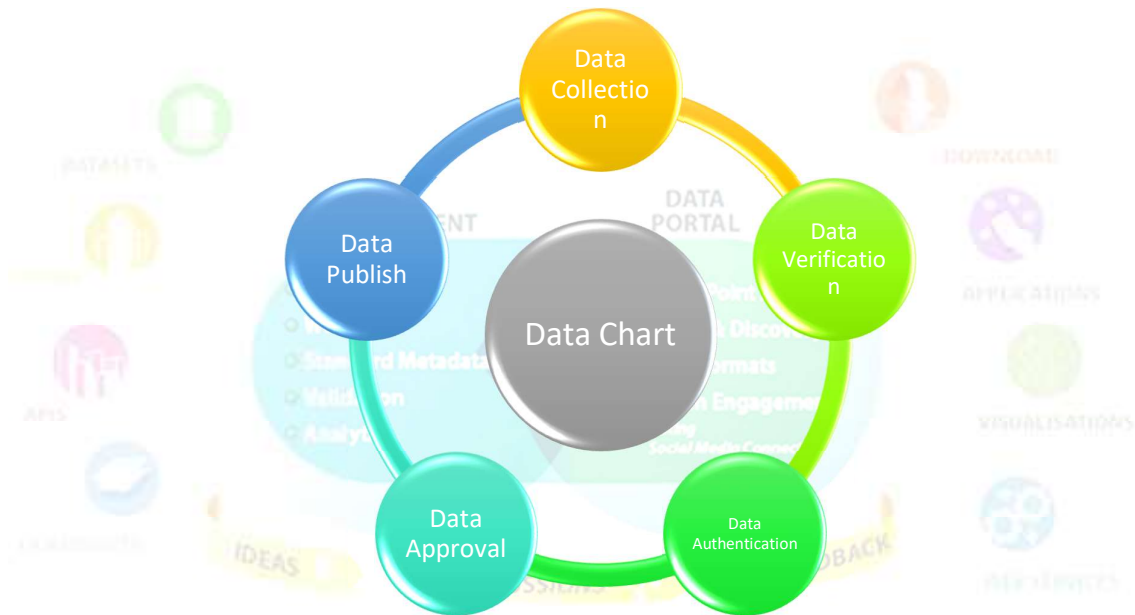
Privacy Policy: Your privacy is important to us. Anyone can generally visit the site without revealing Personal Information unless he chooses to provide such information. We collect and store the following information about user visits for statistical purpose:

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. We make no attempt to link these addresses with the identity of individuals visiting our site unless an attempt to damage the site has been detected.



Data Flow Chart

Agra Municipal Corporation will set up enterprise processes to control the existing available data within the City administration. At every stage of data generation, the concerned stakeholders shall approve and 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. Also, all applications must be developed in compliance with the Government of India's India Enterprise Architecture (IndEA) framework.

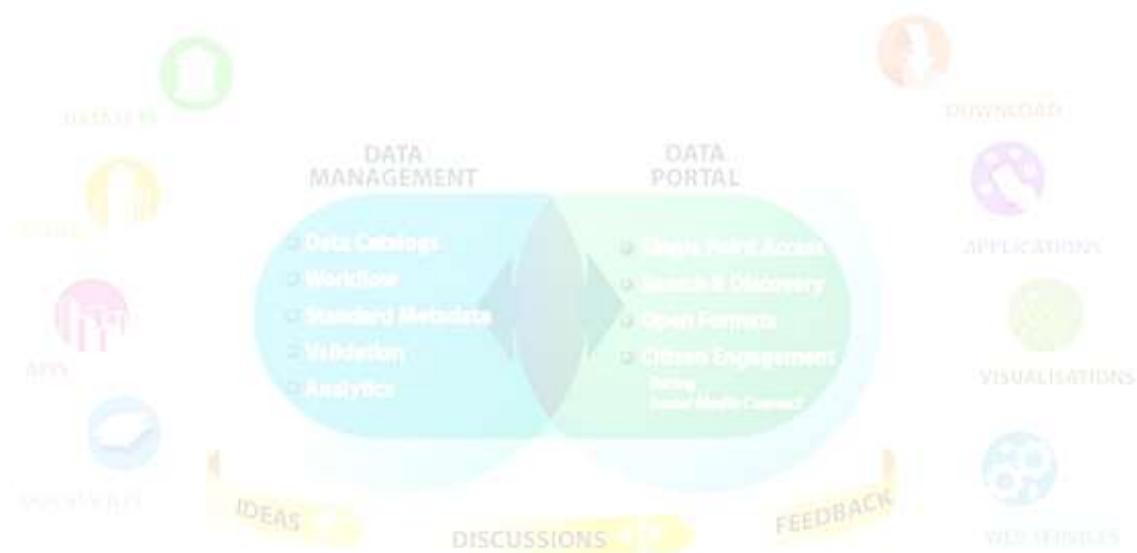


Legal Framework:

Data will remain the property of the agency/ department/ministry/ entity which collected them and reside in their IT enabled facility sharing and providing access. Access to data under this policy will not be in violation of any Act and Rule of Government of India in force. Legal framework of this policy will be aligned with various Acts and rules covering the data.

Pricing:

Pricing of data, if any, would be decided by the data owners and as per government policies. All department will upload the pricing policy of the data under registered and restricted access.



Principals of Open Government Data:

Complete

All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations.

While non-electronic information resources, such as physical artifacts, are not subject to the Open Government Data principles, it is always encouraged that such resources be made available electronically to the extent feasible.

Primary

Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.

If an entity chooses to transform data by aggregation or transcoding for use on an Internet site built for end users, it still has an obligation to make the full-resolution information available in bulk for others to build their own sites with and to preserve the data for posterity.

Timely

Data is made available as quickly as necessary to preserve the value of the data.

Accessible

Data is available to the widest range of users for the widest range of purposes.

Data must be made available on the Internet so as to accommodate the widest practical range of users and uses. This means considering how choices in data preparation and publication affect access to the disabled and how it may impact users of a variety of software and hardware platforms. Data must be published with current industry standard protocols and formats, as well as alternative protocols and formats when industry standards impose burdens on wide reuse of the data.

Data is not accessible if it can be retrieved only through navigating web forms, or if automated tools are not permitted to access it because of a robots.txt file, other policy, or technological restrictions.

Machine processable

Data is reasonably structured to allow automated processing.

The ability for data to be widely used requires that the data be properly encoded. Free-form text is not a substitute for tabular and normalized records. Images of text are not a substitute for the text itself. Sufficient documentation on the data format and meanings of normalized data items must be available to users of the data.

Non-discriminatory

Data is available to anyone, with no requirement of registration.

Anonymous access to the data must be allowed for public data, including access through anonymous proxies. Data should not be hidden behind “walled gardens.”

Non-proprietary

Data is available in a format over which no entity has exclusive control.

Proprietary formats add unnecessary restrictions over who can use the data, how it can be used and shared, and whether the data will be usable in the future. While some proprietary formats are nearly ubiquitous, it is nevertheless not acceptable to use only proprietary formats. Likewise, the relevant non-proprietary formats may not reach a wide audience. In these cases, it may be necessary to make the data available in multiple formats.

License-free

Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.

Because government information is a mix of public records, personal information, copyrighted work, and other non-open data, it is important to be clear about what data is available and what licensing, terms of service, and legal restrictions apply. Data for which no restrictions apply should be marked clearly as being in the public domain.

Online & Free

Information is not meaningfully public if it is not available on the Internet at no charge, or at least no more than the marginal cost of reproduction. It should also be findable.

Permanent

Data should be made available at a stable Internet location indefinitely and in a stable data format for as long as possible.

Designed with Public Input

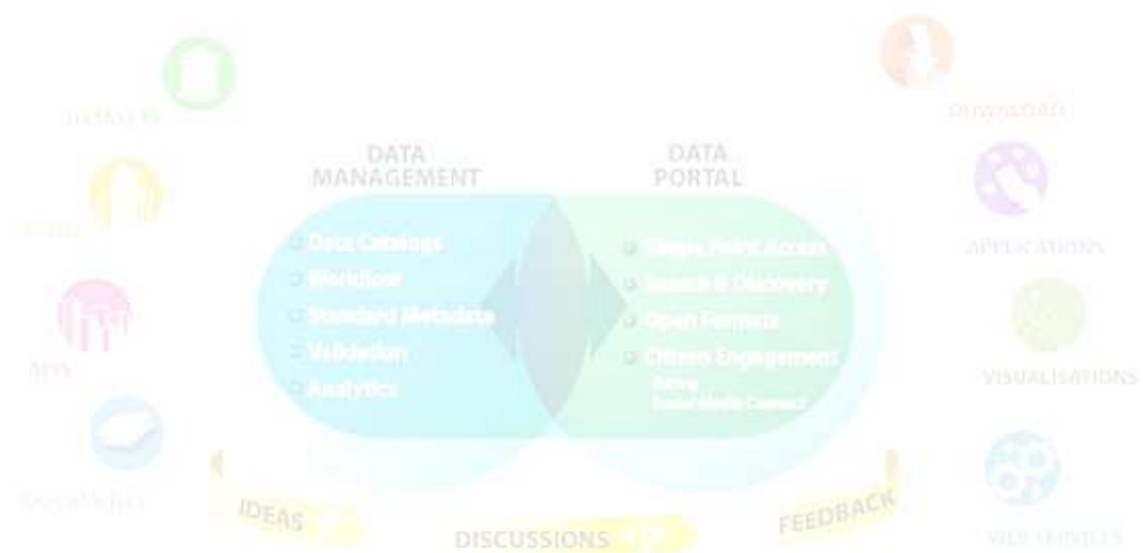
The public is in the best position to determine what information technologies will be best suited for the applications the public intends to create for itself. Public input is therefore crucial to disseminating information in such a way that it has value.

Source: <https://opengovdata.org/>

Salient features:

- Agra Municipal Corporation/Agra Smart City Limited contributes/publishes resources in open format (CSV, XLS, XML, ODS, JASON) either directly or by Web Services through workflow based Data Management System.
- Provides single window access to the datasets and apps published by Agra Municipal Corporation/Agra Smart City Limited of in open format.
- People can directly write to Chief Data Officer (provider) (previously called Data Controller) seeking any further clarification/information on the released resources or related requirements.
- Platform enables better discovery and usage of Government datasets through visualizations and development of apps, mash-ups, etc. Visualization Platform with facility to create maps as well as different chart options like radar, bar, line, area, pie, column, etc.
- Application Programming Interfaces (APIs) to Query Datasets - Direct and dynamic query to access data items of selected datasets have also been provided through the APIs.

- Platform has Responsive Web Layout design i.e. service available on multiple web/mobile platforms.
- People can rate the resources (datasets/apps) on three aspects i.e. Quality, Accessibility and Usability on the scale of 5.



SoP for Data Collection

Data collection is referred to as the method of collecting information in a systematic way. This is the first step towards data-driven decision making and evidence-based governance. Data collection provides both a baseline to measure and a target to improve. In order to get the primary datasets we have to collect data from different wards, zones which is under Agra Municipal Corporation/ Agra Smart City Limited employee and Other Government departments. Data which is not going to be used for any kind of analysis or will not be used for any communication purpose should not be collected at all.

If the request is received from external agency:

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

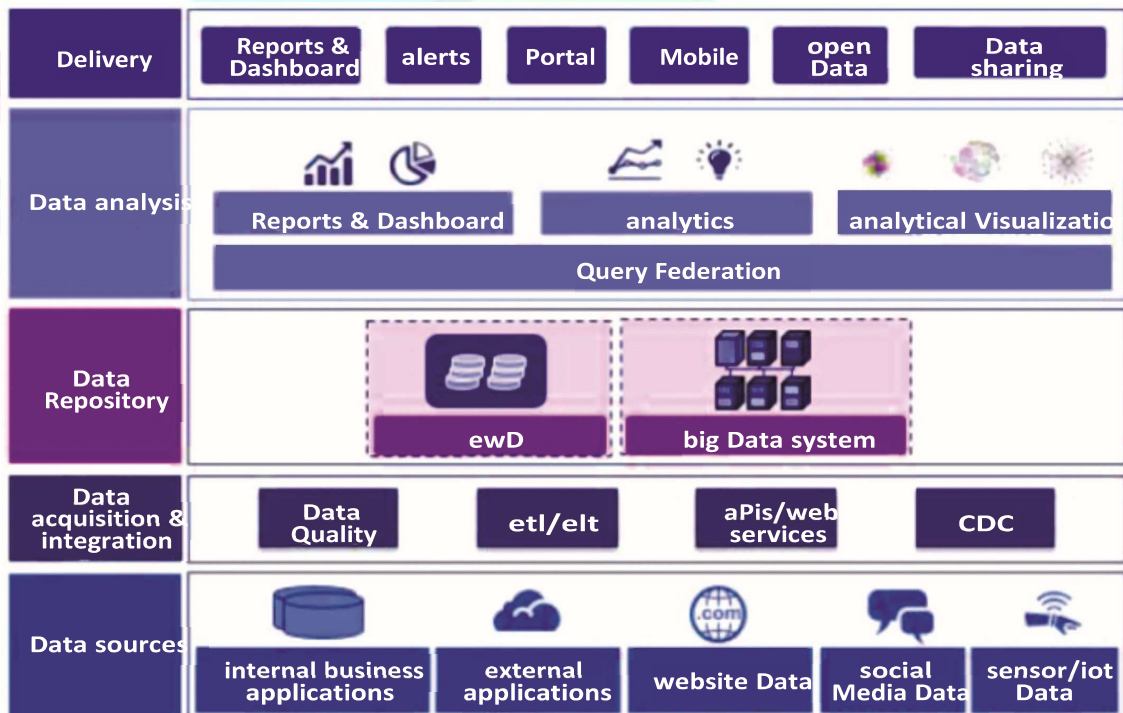
If the request is received from internal departments:

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

SoP for Electronic Data Collection

Cities ecosystem comprises of various technology solutions ranging from Sensors, IoT, SCADA, Electronic camera, GIS, payments system etc. which generate loads of structured data every second on different dimensions. Cities could also leverage various unstructured data emerging from different sources and third-party systems like social media, internet, websites, videos, images etc.

1. The data should be collected with consent of the end-user who may be a citizen or Agra Municipal Corporation/Agra Smart City Limited employee.
2. Special care needs to be taken for data privacy and security. The measures mentioned in above sections of the Policy shall be followed to maintain confidentiality and security of data.
3. For data collection and integration with the IoT systems / sensors, the following Data Architecture highlighted below shall be followed by Agra Municipal Corporation/Agra Smart City Limited.



SoP for Data Processing and Cleaning

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

1. While collecting the electronic data, the IT applications/ IT systems should be developed in such a way that under any circumstances these applications/systems should not accept any wrong data/null data.
2. If there is existing data, identify discrepancies which may come from different sources.
3. The collected data shall be properly processed and cleaned before performing any kind of analysis.
4. If needed, commercial software available in the market can be used with prior approvals from the concerned authorities.

Extract, Transform and Load (ETL) is the common methodology used for data integration and processing. It is a three-step process which used for data integration to blend data from multiple sources. It's often used to build a data warehouse. During this process, data is taken (extracted) from a source system, converted (transformed) into a format that can be analysed, and stored (loaded) into a data warehouse or other system.

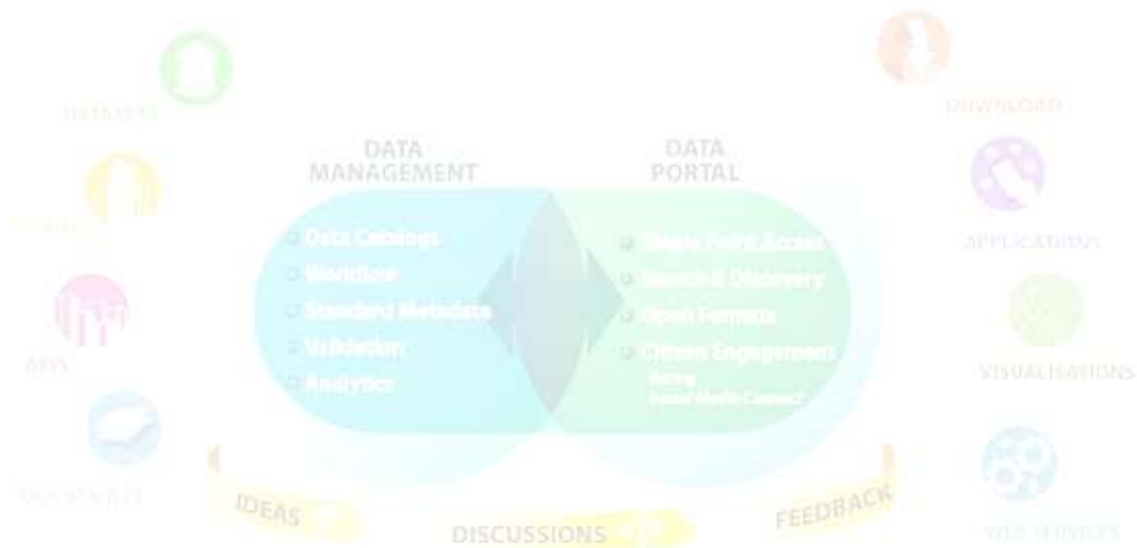
Specifically, Data Champion/ Data Coordinator will keep in mind the following points while collecting data from their respective department:

- a. Spaces in extra columns Compliance
- b. Ward-wise Compliance
- c. Blank Cells Compliance
- d. Standard format Compliance
- e. All NA Compliance
- f. Special Characters Compliance
- g. Split Sheets Compliance
- h. Datasets Compliance
- i. Data Completeness

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.

1. Responsibility for quality check of data rests with the CDO while publishing data on the open data portal or sharing it with stakeholders upon request.
2. Under some circumstances, a special committee under the leadership of Municipal Commissioner (Agra Municipal Corporation), comprising of CDO, and Additional/Assistant Commissioner may be formed for data quality assessment.



SoP for Data Publishing

National Data Sharing and Access Policy (NDSAP) defines standards for publishing datasets and feeds on the open data portal. CDOs must ensure adherence towards defined standards and classification. This SOP describes the steps required for publishing data to the Smart Cities Open Data Portal:

1. **Understand the requirement:** Follow proper procedures to collect the relevant data to be uploaded on the open data portal. Understand the publishing options and the available datasets.
2. **Process the data:** Ensure data is in an appropriate format to be published on the open data portal. It does not contain any personalized information, is open, authenticated and free from defects.
3. **Prepare to deploy/publish data:** Follow procedures specific to the publishing option you have selected and work with the appropriate team (when necessary) to publish your data. NDSAP recommends that datasets should be published in an open format and should be machine readable. Data format can be chosen from the list recommended by NDSAP, highlighted in the Appendix.
4. **Publish metadata:** Follow established metadata procedures as per NDSAP and any other guidelines laid down by the city to publish metadata on the Portal and create linkages between data and metadata.
5. **Obtain approvals and finalise deployment:** Obtain the appropriate management approvals for your data based on your selected data publishing option, from Municipal Commissioner.

SoP for engaging stakeholders to assess the data needs

The concept recognizes the value of enhancing engagement among all four stakeholders of the quadruplehelix model—Government (Agra Municipal Corporation Departments, Agra Smart City Limited and Other Departments of Governments), citizens, academia, and industry, along with improvements in the internal workflow and decision-making processes of city governments.

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

key activities may include:

- a. Identifying stakeholders from various age groups and ethnicities and engaging them in city initiatives
- b. Organising workshops, hackathons/events to promote brainstorming over required datasets
- c. Decision making and consultation with data experts to zero down on the most important datasets required on the portal
- d. Data ideation with public forum to gain understanding of citizen/industry needs
- e. Formation of city data alliances.
- f. Methods used to consult stakeholders include:
 - Phone /email;
 - One-on-one interviews;
 - Workshop/focus group discussions;
 - Distribution of pamphlets and newsletters;
 - Public meetings; and
 - Newspaper/magazines/radio.

SoP for data collection, processing and analysis for on field survey

Field survey is defined as collection and gathering of information at the local level by conducting primary surveys. On-field surveys may be required in situation where data from digital sensors or existing datasets are inadequate. Surveys may be administered to fulfil a certain gap, with a well-defined problem for investigation. Data collected from on-field surveys can help Agra Municipal Corporation/Agra Smart City Limited get a snapshot of how things are at a specific time. The survey research may be descriptive, analytical or evaluative. Field surveys are cost-intensive hence they may be conducted after a thorough mapping of their objectives and outcomes.

survey data processing consists of four important steps:

1. **Survey designing:** Survey is a research strategy and not a method. CDOs/concerned team should choose the most appropriate method based on purpose. CDO shall help the concerned team devise suitable surveys/questionnaires by clarifying the objective, determining sample and deciding upon the survey mode to finally create the questionnaire. A research method should not advise the questions, but other way around. Posing difficult to answer questions, in such case a simple rule or scale can be used to help respondents.
2. **Data collection:** Survey can employ a range of methods including questionnaires, interviews or even focus group discussions. Data entry in either format may happen manually or electronically. 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 intervention of humans. As most of process would be automated and handled by an e-mode, data will be available for further analysis.
3. **Data processing:** Before any analysis is possible, ensuring accuracy and quality of data is paramount. Survey form data is always prone to errors, omissions and other inconsistencies. This data inconsistency and incompleteness, if not edited and corrected on time, can complicate the analysis and may even result in wrong analysis. Data processing shall comprise of various steps necessary for preparing the data for analysis, including editing, data classification, removing redundancies, and

preparation of tables. This is an important step when the survey instrument collects qualitative data, which needs to be then represented in a format for analysis.

4. **Data analysis:** Data analysis covers the final step of characterizing and interpreting research findings. In situations where the digital tools are employed for the survey, and the data can be processed easily. Data analysis will involve computation of certain indices or measures along with searching for patterns of relationship that exist among the data groups. The task of analysing quantitative data may be accomplished through statistics. Descriptive statistics is to be used for organising raw data obtained in the process of research, such as tabulation and classification of data. Inferential statistics, also known as sampling statistics, will be used for making inferences or conclusions from the data collected from a sample and drawing generalisations on the entire population.

