# Provisional City Data Policy for Kanpur Smart City

## Dated: -15th April 2019

# Contents


DATA POLICY FOR "KANPUR SMART CITY ..... 3

1. Need for a Policy ..... 3
2. Introduction ..... 3
3. Objectives ..... 3
4. Scope of this policy ..... 3
5. Appointments ..... 4
6. Formation of City Data Alliances ..... 5

Data Classification ..... 5

DATA CATEGORIZATION ..... 6

Metadata ..... 6

Historical Data ..... 6

Temporary Data ..... 6

7. DATA FLOW/APPROVAL FRAMEWORK ..... 7
8. DATA ARCHIVAL & RETENTION ..... 7
9. DATA SECURITY ..... 7

Software & Hardware-based mechanisms for protecting data ..... 7

Backups ..... 8

Data Masking ..... 8

Data Erasure ..... 8

10. DATA COLLECTION/ELECTRONIC DATA COLLECTION ..... 8
11. DATA ANALYSIS ..... 9
12. Types of Access ..... 9
13. Open Data Norms ..... 9
14. Technology for sharing and access ..... 10
15. Legal framework ..... 10
16. Pricing ..... 10
17. Implementation ..... 11
18. Budget Provisions ..... 11
19. Conclusion ..... 12

# DATA POLICY FOR "KANPUR SMART CITY

## 1. Need for a Policy

There is need for cities to be thoughtful about what data cities collect, why they are collecting it and how they will use it. The City Data Policy has been designed around certain fundamental principles, with the intention of helping cities undergo a process of self-evaluation, of individual goal-setting, and getting themselves ready to embrace a data culture that is suited to their needs and requirements Policy respects that cities have limited bandwidth to address the myriad issues they deal with daily. Therefore, the concept of minimalism has been adopted to ensure that cities collect only that data which can help improve their functioning and provide them with critical insights for the policy-making process. In continuation with directives contained in the NDSAP (National data sharing and accessibility policy), the data policy for Kanpur Smart City has been developed.

## 2. Introduction

Evidence based planning of socio-economic developments processes rely on quality data. There is a general need to facilitate sharing and utilization of the large amount of data generated and residing among the entities of the Government of India. This would call for a policy to leverage these data assets which are disparate. The Current regime of data management does not enable open sharing of Government owned data with other arms of the government nor does it expect proactive disclosure of sharable data available with data owners. Such regimes could lead to duplication of efforts and loss of efficiency of planning of activities focussed on national development. Efficient sharing of data among data owners and inter and intra governmental agencies and with public calls for data standards and interoperable systems. Hence data policy aims to provide an enabling provision and platform for providing proactive and open access to the data generated through public funds available with various departments/organizations of Government of India.

## 3. Objectives

The Objective of this policy is to facilitate the access to Governments of India owned shareable data and information in both human readable and machine-readable forms through "People, Process and Platform". Data Maturity Assessment Framework has been used to drive effective use of data by Kanpur Smart City, and to help the city leaders in structuring their approach to building a collaborative data ecosystem within the Acts and rules of Government of India, thereby permitting wider accessibility and use of public data and information.

## 4. Scope of this policy

The Data Policy of Kanpur Smart City rests on the two most important pillars of the DMAF; Systemic Maturity and Sectoral Maturity. Systemic Maturity comprises Policy, People, Processes, technology and outcomes whereas the Sectoral Maturity is achieved through Data Availability, Data Usage, Data Shareability and Data Management.

This Data **Policy** will assist in data governance, empowerment, protection, collaboration and innovation. It also guides evolution of necessary budgetary allocations to operationalize the policy to use efficiently all data and information created, generated, collected and archived for perspective

planning using public funds provided by Government of India directly or through authorized agencies and autonomous bodies properly. This policy is being issued with the due diligence and approval of the Chief Executive Officer of Kanpur Smart City Limited.

The policy also aims to define the empowered city officials **(People)** with the capacity to guide the development of city data policies, manage data governance, and drive interdepartmental and inter-agency data exchange and to build city data alliances.

The **Processes** involved will assist in assessing the effectiveness of the city's processes around data collection, usage, management, security, privacy, empowerment, collaboration, and innovation. The data collection is being proposed through use of associated data centre of the Integrated Command Control Centre being established in Kanpur.

The **Technology** will assist in assessing the quality and robustness of the city's information and communications technology infrastructure including digital platforms, sensors, IoT devices, data exchanges, big data and artificial intelligence Various platforms are under the process of being integrated with ICCC and proposed ones will be included once they are up and running. The Major Platforms proposed to be integrated are SCADA for electcity and water supply & metering, Smart poles, Solid Waste Management system, Smart Roads, ITMS, Environmental Sensors and Smart Parking.

The **Outcomes** that are expected when the policy is enforced will enable the senior leadership to assess the quality of outcomes around data driven governance, ease of living, ease of doing business, collaboration and innovation in the Kanpur city. The benefits of the data sharing policyare enumerated as under:

a. Maximising use: Ready access to government owned data will enable more extensive use of valuable public resources for the benefit of the community.
b. Avoiding duplication: By sharing data the need for separate bodies to collect the same data will be avoided resulting in significant cost savings in data collection.
c. Maximised integration : By adopting common standards for the collection and transfer of data, integration of individual data sets may be feasible.
d. Ownership Information: The identification of owners for the principal data sets provide information to users to identify those responsible for implementation of prioritized data collection programs and development of data standards.
e. Better Decision – Making: Data and information facilitates making important decisions without incurring repetitive costs. Ready access to existing valuable data is essential for many decision making tasks such as protecting the environment, development planning, managing assets, improving living conditions, national security and controlling disasters.
f. Equity of access: A more open data transfer policy ensures better access to all bonafied users.

With this scope and background the systemic maturity of Kanpur to implement the DataSmart Cities strategy is being issued to achieve the following Aim:

a. Appointment/nomination of key officials with clearly defined roles and responsibilities
b. Formation of City Data Alliances
c. Identification and classification of key municipal data sets
d. Development of a draft city data policy with supporting budgetary allocations

## 5. Appointments

The officers nominated in accordance of the Data Policy being promulgated are described as under:

a. Chairperson: Chief Executive Officer, Kanpur Smart City Limited with the responsibility to review and issue the policy guidelines as per the DMAF guidelines.
b. Additional Chief Executive Officer, Kanpur Smart City Limited to form Data Alliances with other departments that are coexisting and are interdependent on each other.
c. Assistant Commissioner Kanpur Nagar Nigam to coordinate and liaise, procure, update, process the data sets in order to disseminate processed information to senior leadership for taking informed decisions, PPP, budgetary allocations, road map for future development etc.
d. Zonal Officers. The Zonal officers of all six zones will assist the assistant Commissioner in procuring and verifying data being obtained from within their jurisdiction.

## 6. Formation of City Data Alliances

The formation of Data Alliances is mandatory requirement to obtain data required from various agencies and departments which contributing to Kanpur city's development. The major alliance partners are Public Works Department, District Magistrate, Kanpur Nagar Nigam, Police, Jal Kal Vibhag, Kanpur Development Authority, RTO, Jal Nigam, KESCO, AMRUT, UPPCB, Transport Department, Health, Sanitation and Pollution Department to name a few. The data to be obtained from the alliance partners needs to be organised and structured to obtain the required inputs to fulfil the planned road map for development of Kanpur Smart City. The data so acquired in structured form can thus be processed to obtain the right information required for developing a Master Plan, Perspective Procurement Plan, predict and demand budgetary allocations for forthcoming projects in an efficient manner.

The data can be organised and structured as per the Data model enumerated in subsequent Paras.

## Data Classification

Different types of data sets generated both in geospatial and non-spatial form by different departments and autonomous bodies functioning in Kanpur are to be classified as:-

a. **Sharable data** –The Data not covered under the scope of negative list and non-sensitive in nature.
b. **Negative List**- Non-Sharable data as declared by departments/organizations. This could be confidential data related to pricing /tender cost etc
c. **Restricted Data** –Data which are accessible through a prescribed process of registration and authorization by respective departments/ organizations.
d. **Sensitive Data** – The data not covered under the scope of negative list and non-sensitive in nature.

The types of data produced by a statistical system consist of derived statistics like national accounts statistics, indicators like price index, databases from census and surveys. The geospatial data however, consists primarily of satellite data, maps, etc. In such a system, it becomes, important to maintain standards in respect of metadata, data layout and data access policy. All departments/ministries will prepare the negative list within six months of the notification of the policy, which will be periodically reviewed by the oversight committee.

## DATA CATEGORIZATION

a. **Master Data** - Master data describe about the people, places, and things that are involved in.. Examples include people (e.g., customers, employees, vendors, suppliers), places (e.g., locations, sales territories, offices), and things (e.g., accounts, products, assets, document sets). Because these data tend to be used by multiple trades, standardizing master data formats and synchronizing values are critical for successful system integration. Master data tend to be grouped into master records, which may include associated reference data.

b. **Transactional Data** -Transactional data describe an internal or external event or transaction that takes place as an organization/departments conducts its business. These data are typically grouped into transactional records, which include associated master and reference data.

c. **Reference Data** - Reference data are sets of values or classification schemas that are referred to by systems, applications, data stores, processes, and reports, as well as by transactional and master records. Standardized reference data are key to data integration and interoperability and facilitate the sharing and reporting of information. Reference data may be used to differentiate one type of record from another for categorization and analysis, or they may be a significant fact such as country, which appears within a larger information set such as address. Organizations often create internal reference data to characterize or standardize their own information. Reference data sets are also defined by external groups, such as government or regulatory bodies, to be used by multiple organizations.

## Metadata

Metadata literally means "data about data." Metadata label, describe, or characterize other data and make it easier to retrieve, interpret, or use information. Technical metadata are metadata used to describe technology and data structures. Business metadata describe the nontechnical aspects of data and their usage. Audit trail metadata are a specific type of metadata, typically stored in a record and protected from alteration, that capture how, when, and by whom the data were created, accessed, updated, or deleted. Audit trail metadata are used for security, compliance, or forensic purposes. Although audit trail are typically stored in a record, technical metadata and business metadata are usually stored separately from the data they describe. These are the most common types of metadata, but it could be argued that there are other types of metadata that make it easier to retrieve, interpret, or use information. The label for any metadata may not be as important as the fact that it is being deliberately used to support data goals. Any discipline or activity that uses data is likely to have associated metadata.

## Historical Data

Historical data contain significant facts, as of a certain point in time, which should not be altered except to correct an error. They are important to security and compliance. Operational systems can also contain history tables for reporting or analysis purposes.

## Temporary Data

Temporary data are kept in memory to speed up processing. They are not viewed by humans and are used for technical purposes.

## 7. DATA FLOW/APPROVAL FRAMEWORK

The data flow could be organised through regular report and returns to be obtained from various department functioning in the Kanpur Smart city through regular reports and returns to be submitted on a weekly/fortnightly basis depending on the criticality of the data. This data can be initially obtained in hard copies in standard formats as finalised by the officers nominated for coordinating and procurement of data from various sources. Parallely the nominated officers will work towards creation of Database for storing and accessing the data digitally through API queries in SQL database servers with required data application support as part of Kanpur Smart City initiative.

The approval process will be followed as per the guidelines issued in DMAF by MoHUA by the officers nominated/appointed for this purpose as mentioned in this policy. The approval process will involve the analysis and verification of data as per the type of data structured described in this policy.

## 8. DATA ARCHIVAL & RETENTION

Archiving is the process of moving data that is no longer actively used to a separate storage device for long-term retention. Archive data consists of older data that is still important to the organization and may be needed for future reference, as well as data that must be retained for regulatory compliance.

Data archival and retention will be done as per the existing policies on the subject. The data will be archived every month or as per the periodicity decided by the Data Administrator appointed from the IT department of Kanpur Nagar Nigam/Kanpur Smart City Limited by the Chairman. The archived data can be stored preferably in digital format or in hard copies. The use of DVDs, external hard drives or tape libraries can be utilised for storage of archived data. The Document Management System installed as part if the ICCC project may be utilised for converting the archived data into digital format.

## 9. DATASECURITY

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

### Software & Hardware-based mechanisms for protecting data

Software-based security solutions encrypt the data to protect it from theft. However, a malicious program or a hacker could corrupt the data in order to make it unrecoverable, making the system unusable. Hardware-based security solutions can prevent read and write access to data and hence offer very strong protection against tampering and unauthorized access.

Hardware based security or assisted computer security offers an alternative to software-only computer security. Security tokens such as those using may be more secure due to the physical access required in order to be compromised. Access is enabled only when the token is connected and correct PIN is entered (see two-factor authentication). However, dongles can be used by anyone who can gain

physical access to it. A newer technology in hardware-based security solves this problem offering full proof security for data.

Working of hardware-based security: A hardware device allows a user to log in, log out and set different privilege levels by doing manual actions. The device uses biometric technology to prevent malicious users from logging in, logging out, and changing privilege levels. The current state of a user of the device is read by controllers in peripheral devices such as hard disks. Illegal access by a malicious user or a malicious program is interrupted based on the current state of a user by hard disk and DVD controllers making illegal access to data impossible. Hardware-based access control is more secure than protection provided by the operating systems as operating systems are vulnerable to malicious attacks by viruses and hackers. The data on hard disks can be corrupted after a malicious access is obtained. With hardware-based protection, software cannot manipulate the user privilege levels. It is impossible for a hacker or malicious programs to gain access to secure data protected by hardware or perform unauthorized privileged operations. This assumption is broken only if the hardware itself is malicious or contains a backdoor. The hardware protects the operating system image and file system privileges from being tampered. Therefore, a completely secure system can be created using a combination of hardware-based security and secure system administration policies.

### Backups

Backups are used to ensure data which is lost can be recovered from another source. It is considered essential to keep a backup of any data in most industries and the process is recommended for any files of importance to a user.

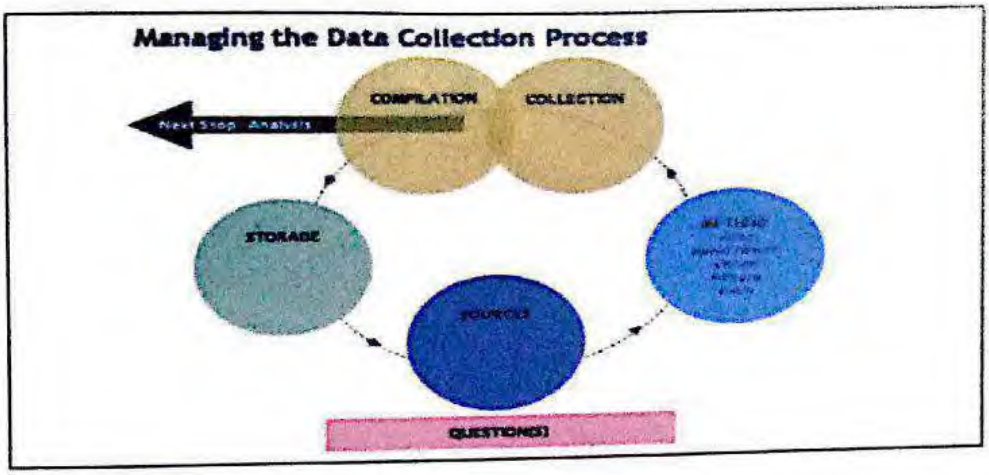
### Data Masking

Data masking of structured data is the process of obscuring (masking) specific data within a database table or cell to ensure that data security is maintained and sensitive information is not exposed to unauthorized personnel. This may include masking the data from users (for example so banking customer representatives can only see the last 4 digits of a customer's national identity number), developers (who need real production data to test new software releases but should not be able to see sensitive financial data), outsourcing vendors, etc.

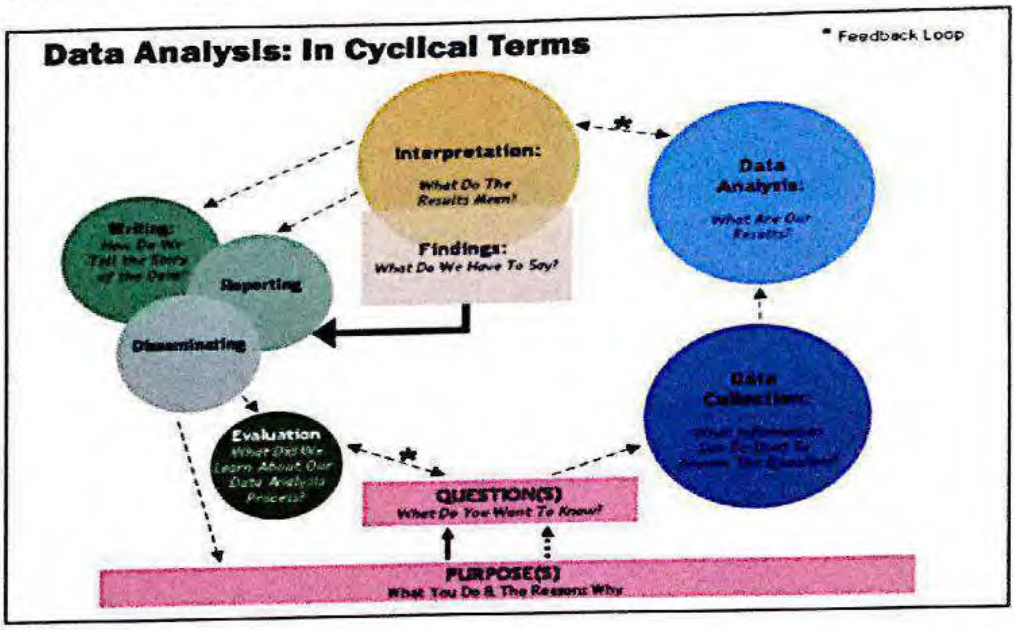
### Data Erasure

Data erasure is a method of software based overwriting that completely destroys all electronic data residing on a hard drive or other digital media to ensure that no sensitive data is lost when an asset is retired or reused.

## 10.DATA COLLECTION/ELECTRONIC DATA COLLECTION




## 11.DATA ANALYSIS




## 12.Types ofAccess

1. **Open Access:** Access to data generated from public funding should by easy, timely, user friendly and web based without any process of registration/authorization.
2. **Registered Access:** Data sets which are accessible only through a prescribed process of registration/authorization by respective departments/organizations will be available to the recognized institution/organizations/public users through defined procedures.
3. **Restricted Access:** Data declared as restricted, by government of India policies, will be accessible only through and under authorization.

## 13.Open Data Norms

Data are "open" when they are always published and updated online as soon and as often as possible, in a way that allows, at the lowest possible cost, to legally reuse them for free, for any purpose (including for-profit activities!) and to quick and easy automatically process them with any software. In practice, raw data are open when they have an open access license that allows what described in the previous sentence and are published in an open file format, or are directly accessible with open protocols not hindered by patents or similar restrictions, through the Internet.

## 14.Technology for sharing and access

A state of the art data warehouse and data archive with online analytical processing (OLAP) capabilities, which includes providing, a multi-dimensional and subject oriented view of the database needs to be created. This integrated repository of data portals of various ministries/departments as a part of data.gov.in will hold data and this repository over a period of time will also encompass data generated by various state governments of UTs. The main features of the data warehouse need to include:

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

## 15.Legal framework

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

## 16.Pricing

Pricing of data, if any, would be decided by the data owners and as per the government policies. All ministries/Department will upload the pricing policy of the data under registered and restricted access within three months of the notification of the policy. A broad set of parameters would be standardized and provided as guidelines for the use of data owners.

## 17. Implementation

a. The Kanpur Smart City (KSCL) will perform the nodal functions of coordination and monitoring of policy through close collaboration with all the department functioning under the Divisional Commissioner of Kanpur.
b. All Sharable data will be made available on as is where is basis.
c. Detailed implementation guidelines including the technology and standards for data and metadata would followed as per the policies laid down by Department of Information technology, Government of India.
d. All the data users who are accessing/using the data shall acknowledge the KSCL/department in all forms of publications.
e. All departments will upload/ provide at least 5 high value data sets initially through email or through APIs to the central data server proposed to be set up in ICCC. Later on the data will be updated by KSCL on data.gov.in within three months of the notification of the policy.
f. Uploading of all remaining data sets should be completed within one year.
g. Thereafter, all data sets are to be uploaded regularly every quarter.
h. Data.gov.in will have the metadata and data itself and will be accessed from the portals of the departments/ministries.
i. The metadata in standardized formats is to be ported by KSCL on data.gov.in which enables data discovery and access through departmental portals. All metadata will follow standards and will minimally contain adequate information on proper citation, access, contact information, and discovery complete information including methods, structure, semantics, and quality control/assurance is expected for most datasets.
j. KSCL will design and position a suitable budgetary incentive system for data owners for increasing open access to the sharable data.
k. An oversight committee will be constituted by Chairman, KSCL for facilitating the implementation of the policy and its provisions thereof. This oversight committee will work under the co-ordination committee for implementation constituted by Department of Information Technology

## 18. Budget Provisions

The data policy is expected to entail expenditure for both data owners and data mergers for analog to digital conversion, data refinement, data storage, quality up-gradation etc. Budgetary provisions and appropriate support for data management for each department/organization would be necessary.

## 19.Conclusion

While policies provide official mandate, facilitation of optimum accessibility and usability of data by the implementers pre suppose a trajectory of proper organization of data, with access services and analysis tools that provide the researchers and stakeholders with added value. For data to be reused, it needs to be adequately described and linked to services that disseminate the data to other researchers and stakeholders. The current methods of storing data are as diverse as the disciplines that generate it. It is necessary to develop institutional repositories, data centres on domain and national levels that all methods of storing and sharing have to exist within the specific infrastructure to enable all users to access and use it.

The Policy aims at the promotion of a technology based culture of data management as well as data sharing and access. It opens up, proactively, information on available data, which could be shared with civil society for developmental purposes, their price details if any, and methods for gaining access to registered and restricted sue. The policy has limited its scope to data owned by the agencies, departments/ministries and entities under the government of India and forms a statement of government of India of its commitment to transparency and efficiency in governance. Department of Science and Technology will continue the process of evolving the policy.

(Municipal Commissioner)