KARNATAKA MUNICIPAL OPEN DATA POLICY

2020

KARNATAKA URBAN INFRASTRUCUTRE FINANCIAL CORPORATION

GOVERNMENT OF KARNATAKA

ABBREVIATIONS

DPA - Data Protection Authority

KMA – Karnataka Municipalities Act

KMC – Karnataka Municipal Corporation

KMDS – Karnataka Municipal Data Society

NDSAP - National Data Sharing and Accessibility Policy

ULB's – Urban Local Bodies

MoHUA - Ministry of Housing and Urban Affairs

TABLE OF CONTENTS

OBJECTIVE OF KARNATAKA MUNICIPAL OPEN DATA POLICY

VISION

Access to information is an important determiner of the success of a Government's relationship with its citizens. Consequently, the ethos behind open data emanates from the notion that data is a public asset and hence, should be in an open and usable format which can bring value to all stakeholders in an ecosystem. As the value of data cannot be predicted, it is essential in today's world to have an **open by default policy** for as much information as possible. Having limited legal, financial or technological restrictions will ensure the potential of the public data to be realized by all stakeholders. Traditionally, it is the Government that is a repository of most information on citizens and the services it provides, which if harnessed would add significant while formulating decisions.

For the benefits of government data to be maximised it should be in a format that enables third parties to re-use and create value from that data, as well. The primary purpose behind an open data policy in this context is then, to look at how the range of information published by urban local bodies can be expanded and made as useful as possible to citizens, business, the voluntary sector and government itself. To that effect we propose the following policy statements,

The vision of the policy is threefold;

1. Improve public understanding of government functioning by increasing transparency, accountability and awareness about their public services and resources.

2. Facilitate innovation by providing data points to locate problems and facilitate solutions by individuals and partnerships.

3. Identify policy trends by analysing and reviewing the data points.

PRINCIPLES

Government of Karnataka through this policy recognizes the principles set forth in The International Open Data Charter. The charter was developed in 2015 by governments, civil society, and experts around the world to represent a globally-agreed set of aspirational norms for how to publish data. Karnataka Municipal Data Policy, on the lines of the International Open Data Charter, has identified four key principles it will be guided by-

1. Open by default: Having a policy *that presumes the publication of all data is the first step towards a good open data policy.* It ensures that any data that is then kept closed has to be justified on the basis of established legal exceptions in terms of national security or data protection.

2. Timely and Comprehensive: Having *up to date information is vital for a data initiative to be relevant*, along with maintaining its authenticity in terms of publishing it in its original and unmodified form.

3. Accessible and Usable: The *format of the data is vital in its accessibility such as it being machine readable, easily discoverable, and for usability* it should be under an open license such as one under Creative Commons.

4. Comparable and Interoperable: Technical standards such as commonly-agreed upon data standards are key *in order for collaboration of datasets and maximizing their value.*

SCOPE OF THE POLICY

The policy is applicable for all the data generated by virtue of the functions granted by the Karnataka government to its urban local bodies under

a) Section 58 and 59 of the Karnataka Municipal Corporations Act, 1976.

b) Section 87, 88 and 91 of the Karnataka Municipalities Act, 1964

Since this policy on municipal data is the first effort towards providing a structural framework for opening and releasing data, Government of Karnataka extends this policy to all smart cities identified currently and under the Government of India's Smart Cities Mission in Karnataka. It would also stand extended to those cities which may be covered in the future phases of the Smart Cities Mission.

[KUIDFC TO RE-CONFIRM]

RE-IMAGINING URBAN DATA

Municipal bodies, across the country, manage large quantities of data pertaining to issues of urban governance such as water management, waste management, weather management, housing and such. Targeted solutions to many issues in those aspects can be driven through an approach revolving around analysis of the current data, to ascertain future trends. However, it is not only government agencies but also private entities, civil society, research and academic organizations which contribute to urban functioning by interacting with the municipal data and creating more value for such data. The variety in datasets created by the different entities, if utilized efficiently, can create an ecosystem of data that feeds into the best efforts of each stakeholder.

Such data can enhance governance by improving decision making and aiding in formulation of policies. This necessitates having accurate data that can be analysed in its current format at the precise time needed, to indicate how civic bodies are managing citizen needs such as dealing with consumer complaints. Eventually, such information across an area can be used to assess the trends, citizen satisfaction ratings and patterns to allocate resources across various departments based on data.

The benefits of open data can be observed in various spheres of governance such as enhanced transparency and accountability of how public funds are utilized. When municipal information is made open, it gives greater insight into the functioning of the bodies itself with a larger number of entities having access to information that can provide valuable insights. Feedback from citizens and other stakeholders can then aid in increased efficiency of public services, as well. Additionally, while the Right to Information(RTI) Act empowers citizens in seeking information for their specific queries, open data can help reduce such enquiries from the government by proactively publishing data. The availability of such information can also aid in government officials responding to concerns of citizens better, since they are armed with more public information. Lastly, there is reason to believe that the utilization of open data by different commercial and non-commercial entities can lead to development of new businesses and services. Interplay of open data between different stakeholders can create synergies beneficial to our society.

Conventionally, open data revolves around collection of data about places and organization as opposed to focusing on individuals. Currently, the Karnataka Municipal Data Society (KMDS), through its website, releases data from ULB's on the finances of the area, elected representatives, municipal functions such as accounts, birth and deaths, gas connections, trade licenses, consumer complaints, water connections etc. Urban Local Bodies are expected to keep information on municipal services they undertake such as the above and this is released on their individual ULB website which can be accessed through the KMDS website, as well.

LEGAL GUIDANCE

Open Government Data is mainly concerned with sharing of non-personal data. The Data Protection Bill governs usage of personal data and hence even when it is passed, it will have limited relevance. The exception being Clause 91 (1) of the Bill which allows the Central Government to collect anonymized or non-personal data from any data fiduciary for improvement of services or usage in evidence-based policy making.

Karnataka can make laws given the right of a State under Article 243W of the Constitution to legislate the powers, authorities and responsibilities of its Municipalities. Entry 5 of List II also gives a State the power to legislate its local government in terms of the powers of municipal corporations for the purpose of local self-government or administration.

In the case of Justice K.S Puttuswamy(Retd) v. Government of India, which declared Privacy as a fundamental right available to citizens, was centred around personal data. However, it noted that since we are living in the age of big data and to minimize privacy risks in such a period, certain considerations were important to be mindful of. Firstly, digital platforms are imperative in ensuring good governance and thus, states can have justifiable reasons for the collection and storage of data. Secondly, privacy can be a legitimate expectation even when an individual is in a public space. An individual can claim this right of privacy if they establish that there exists a real concern of impending harm due to a violation of their privacy. Such a claim can be contemplated as an infringement under Article 21, the Right to Life after deliberating on the following factors; The context in which a privacy slaim is set up, whether the claim relates to a private or family, or a confidential relationship. It has to be a serious claim, along with the nature and extent of disclosure likely to result in a serious or significant injury. The nature of information, whether already in public domain or whether its personal and sensitive information of an identified individual will also be key in determination of a successful claim.

National Data Sharing and Accessibility Policy (NDSAP)

The National Data Sharing and Accessibility Policy (NDSAP) created in 2012 stands as the formal policy of the central Government. It was implemented to encourage and govern publication of government created and owned datasets, from various agencies of the central government in open digital formats and through a single national data portal. It is applicable to all "shareable non-sensitive data available either in digital or analog forms but generated using public funds by various ministries, departments, subordinate offices, organizations, and agencies of Government of India as well as of the states." "Shareable" refers to data declared to be such by the government agency that created it, and "non-sensitive" refers to data sharing of which is not prohibited by any central government acts. The format of this data is also specified to be in standardized human and machine-readable formats. The policy traces its origins to Section 4 of the Right to Information (RTI) Act which requires the government to proactively publish data collected with the aid of taxpayer money. However, the policy is not binding.

INTERNATIONAL BEST PRACTICES

a) Collect or create information in a way that supports downstream information processing and dissemination activities[1] - The United States White House policy requirements emphasize that every agency handling data must consider how their handling of data at every stage will impact it in the other stages of the information cycle. This is to prevent the data from having to be repurposed at a different stage which can be ensured if all decisions pertaining to its usage are considered and consulted with the relevant parties, at the very first instance. Such as, determining the format and frequency of update.

b) Communicate the value of open data initiatives to internal stakeholders – it has been globally recognized that not only do the benefits of open data have to be highlighted to the public but equally importantly, to the internal management. This is aided by a clear demarcation of roles and responsibilities that each member of the ULB, potentially, has in the process. France, in its reference documentation to its open data roadmap, was keen to explaining the importance of open data to public employees and promoting its adoption.[2]

c) Incorporating Privacy Impact Assessments for each stage of data management: The United Kingdom in its whitepaper on open data highlights the need for an approach where privacy risks are managed subsequent to the completion of the data cycle, considering it during each stage is more efficient and cost effective for the ULB's. Firstly, this is an important precedent and standard to set in government policy making and secondly, it helps in determining the level of risk the organization is being exposed to and the kind of restrictions needed for access to data in order to mitigate the risks.

d) Emphasize on review mechanisms of the policy as much as the policy itself: within a period fixed by KUIDFC, all the ULB's should review and revise, where needed, their data management practices and systems. Alternatively, as South Korea has done, there could be the creation of an Open Data Strategic Council within KUIDFC comprising of the highest officials of the ULB.[3] This differs from the Urban Data Observatory in that it could be of a higher-level discussion on strategies for the progress and new directions that Karnataka's open data efforts can take.

e) Lastly, ULB's or KUIDFC on a larger scale can introduce courses and trainings for the public on the use and re-use of data. Such initiatives have been implemented in the UK and contribute to capacity building on the use of open data. [4]

---

[1] https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-13.pdf

[2] https://www.modernisation.gouv.fr/home/lopen-data-a-son-vade-mecum

[3] http://www.odsc.go.kr/user/nd73172.do

[4] https://datos.gob.es/sites/default/files/doc/file/international_open_data_best_practices.pdf

DATA COLLECTION FRAMEWORK

The next few sections of the policy are vital since the processes that frame the data ensure its legitimacy and validity. The primary intention of this policy is to ensure that all kinds of data being captured by ULB's on ground that can be made public, is available to the citizens digitally. This is possible only when there are appropriate modes of gathering the data collected by the ULB's. Such urban local bodies execute and manage several functions under them that generate data on aspects such as water supply, building plan approvals etc. The data by virtue of their functions is specified in the KMA/KMC Act. Another source of data is through digital tools that exist across the civic infrastructure managed by ULB's such as CCTV cameras or other sensory objects. Specific government programs or schemes, if handled by local municipal bodies, also generate data. It is not only external data but even administrative internal data such as consumer complaints, being handled by the ULB's that can potentially be of significance to stakeholders and hence, is of value opening up. Data collection can be done daily, weekly or monthly depending on the data set.

What data should be open?

The NDSAP lays down the framework for opening data by segregating it into two possible types; a negative list and a high value data list. Wherein; negative list is a set of datasets that should be confidential; which either contain personally identifiable information or concern national security. A High Value Data set is a dataset that is; non-discriminatory, machine readable and easy to access, complete, primary and timeless. For example, in the above-mentioned example of releasing data on consumer complaints, the nature of complaints and steps taken towards its resolution when in an accessible format would qualify as a high value data set. If the dataset contained the names, numbers or any other personal information about the citizens making the complaints then it could fall under the bracket of a negative list.

Three guiding steps towards deciding which data should be made open can be:

1. If the data is commonly sought under the Right to Information Act such as annual report accounts of revenue and expenditure of ULB's

2. If the data increases the knowledge, accountability, efficiency or delivery of the ULB's services, for example -

3. If the opening of the data would create an economic opportunity such as opening up citizen data on

Prioritization of datasets for release

At any given time, there are multiple sources of data being generated by citizen services or municipal work, to differing extents of depth. Yet, not all data are equally significant from the perspective of public consumption. Thus, having a non-exhaustive list such as below can help authorities in discerning which data deserves more resources and priority in terms of being released. It enables quicker decision making and lesser bureaucratic hurdles in authorizing releasing of data if ULB's are aware of the criterion themselves. Placing such responsibility to prioritize data can also help empower and motivate the officials of such municipal bodies.

The Ministry of Housing and Urban Affairs (MoHUA) in its report on the Smart Cities Mission identifies the following factors in prioritizing data feeds:

a) Data sets which promote co-creation and open innovation.

b) Information which is frequently requested by stakeholders.

c) Information which leads to academic research on civic issues.

d) Information which helps in improving operational efficiency of municipal administration.

e) Information which leads to enhanced transparency and accountability.

f) Information which leads to citizen convenience.

g) Information which promotes multidisciplinary research using insights or analysis.

h) Information which leads to institutionalization of a culture of data driven governance

i) Information which leads to better situational awareness.

j) Information which leads to mandatory disclosures.

k) Information which forms the basis of reports/research funded by public funds by city administration.

l) Information which improves inter-department coordination.

Eventually, what needs to be kept in mind is also the purpose of the dataset. The right kind of data will enable stakeholders to make one of the following four analysis:

1) Descriptive analysis – to answer the question of what is happening using the data points

2) Diagnostic analysis – answers the question of why a phenomenon is happening using the data points

3) Predictive analysis – where the data points help ascertain a future trend

4) Prescriptive analysis – where the data points help in answering the question of what can be done to mitigate an issue

The potential of one of the above combined with identifying the purpose of data in terms of whether it is intended for reporting of information, administrative work or planning for the future, gives a fair guide of which datasets should be prioritized.

Data formats

Formats of open data that are most useful are ones that allow for maximum re-usage whether in terms of being released in open standards or machine-readable formats. The Sunlight Foundation

defines open standards as those that store information in a manner which can be accessed by proprietary or non-proprietary software. Machine-readability, on the hand, refers to a data structure understandable by a computer. Developing on those are then machine-processable formats which essentially make it easier for data to found by machine searching and other sorting processes. Conventional formats such as HTML and PDF, pose a higher inconvenience when it comes to converting the information in them whereas data in formats like XML open themselves to easier analysis. Thus, not only do the formats have to be most technically appropriate for further conversion but also most accessible in terms of the user's convenience.

The following file formats are suitable for the ULB's to publish their data since they are prescribed by the Central Government in NDSAP and adhering to the national standard on formats will ensure interoperability of data. Interoperability, as defined in the Open Data Handbook, is the ability of diverse systems such as different datasets in an open data ecosystem, to interact with each other in a bid to build larger, complex systems of data. The formats suggested are:

1. CSV (Comma separated values)

2. XLS (Spread sheet -Excel)

3. ODS/OTS (Open Document Formats for Spreadsheets)

4. XML (Extensive Markup Language)

5. RDF (Resources Description Framework)

6. KML (Keyhole Markup Language used for Maps)

7. GML (Geography Markup Language)

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

Labelling of data

Metadata is described as the data fields contained in a given data set. Each dataset on the city's web portal should have accompanying metadata to inform potential users of the dataset itself.

Good metadata discipline ensures that the data assets are most useable, searchable, and accessible for the public and the government alike. All the metadata should be in standardized formats and at the minimum contain adequate information on proper citation, access, contact information, and discovery. Additionally, it is beneficial to have the frequency of data update along with the time and date of the most recent update.

Publication of data

Published datasets shall be placed into the public domain, in a manner that there are no restrictions or requirements placed on use of these datasets. This is currently being done on the KMDS website but we also suggest creating a new platform for the state's Open Data Policy. Each published dataset should be associated with contact information for the appropriate manager of that dataset as well as with a file layout or data dictionary that provides information about field labels and values. Urban

local departments will specify a recommended data citation form available for viewing on the central online location for published Urban local bodies data to encourage responsible reuse of Urban local bodies data.

OPENING UP DATA

How will the data be made accessible to the public?

The principle behind an open data policy is the maximum publication of municipal data to serve public interest, however, we recognize that this cannot always be accomplished. Urban Local Bodies, often, manage sensitive information of citizens and hence exceptions for security and any other compulsions of municipal governments will need to be carved out. To ensure that ULB's can determine this on a case to case basis, we recommend the following procedure of access:

Open Access – No process of registration or authorization will be needed to access such data that is generated from public funding.

Registered Access – Data sets which are accessible only through a prescribed process of registration or authorization by respective department/organizations and will be available to recognized institutions/organization/public users through defined procedures.

Restricted Access: Data declared as restricted, by Government of India policies, will be accessible only through and under authorization.

[KUIDFC to determine on kinds of data which should be open or restricted]

DATA FOR BETTER GOVERNANCE

a) INTERNAL STRUCTURE

To guide and amplify open data efforts, it is imperative to have a state data policy which outlines the responsibilities and expectations from such an ecosystem. This ensures the proper management of information while lack of one would inhibit systematic progress towards the vision of the open data efforts. Currently, KMDS operates as the organization in charge of releasing open data, however, there is no mandate guiding the framework of such data. The working of the current process is also opaque in terms of a lacunae in how they go about collecting data apart from the information collected by each ULB. Their website links to the website of each ULB where you can find information about the municipal area and avail citizen services however much of the data is not up to date. A framework covering their roles and responsibilities along with those of the ULB's is detailed below:
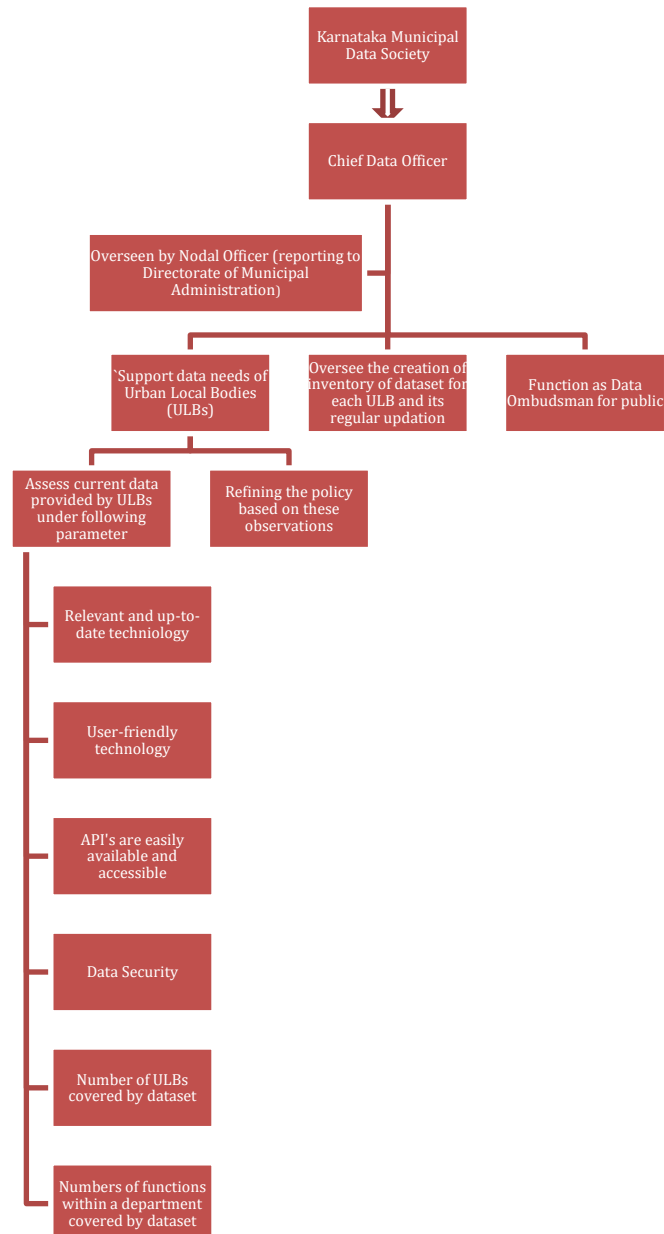
Municipal level
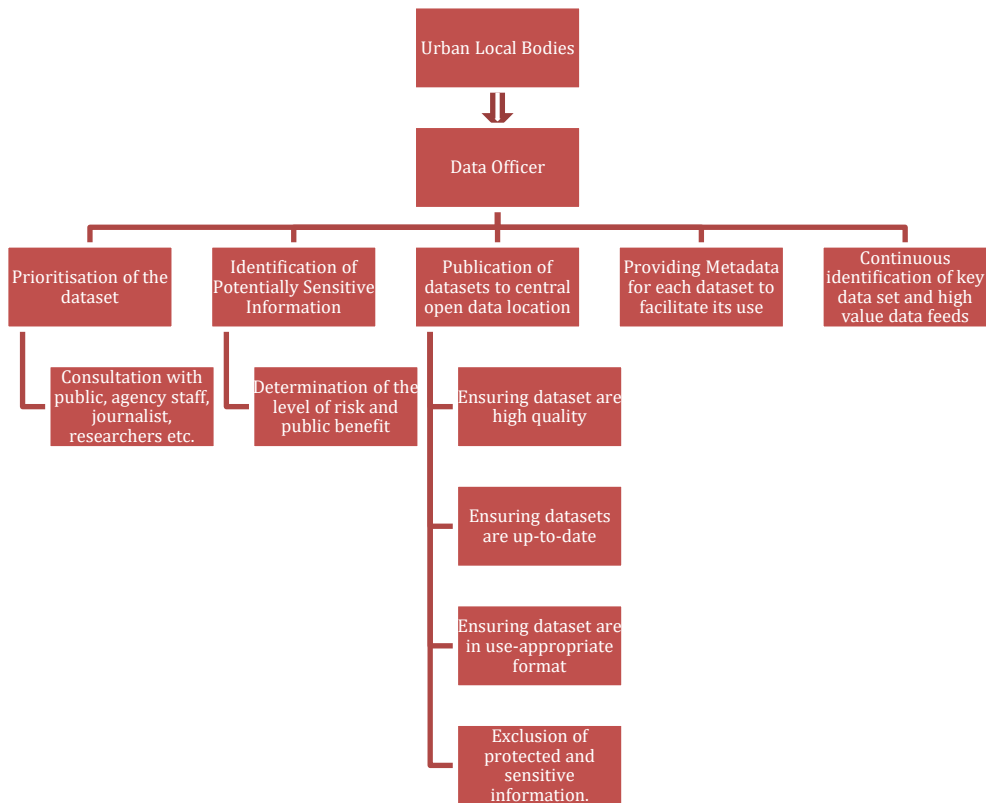
In each ULB, a data officer should be appointed who will essentially be in charge of implementing this open data policy in their respective city. This begins with determining the level of risk and public benefit associated with potentially sensitive, non-protected information so as to decide about

whether and how to publish it. Efforts to guide the prioritization of the datasets as per the criteria established will be supervised by them. The officer will proactively consult with members of municipality such as the public, agency staff, journalists, researchers, and other stakeholders to identify the datasets which will have the greatest benefit to Urban local bodies residents if published in a high-quality manner. He/she will be in charge of overlooking the publication of datasets to the central open data location, including processes for ensuring that datasets are high quality, up-to-date, are in use-appropriate formats, and exclude protected and sensitive information. This includes ensuring that appropriate metadata is provided for each dataset in order to facilitate its use. Lastly, they will ensure that there is a continuous identification of key data sets and high value data feeds to be published on the data Portal. Such lists should be prepared and continuously reviewed in line with NDSAP and other government policies.

Karnataka Municipal Data Society (KMDS)

Karnataka Municipal Data Society
⇓
Chief Data Officer

Overseen by Nodal Officer (reporting to Directorate of Municipal Administration)

`Support data needs of Urban Local Bodies (ULBs)

Oversee the creation of inventory of dataset for each ULB and its regular updation

Function as Data Ombudsman for public

Assess current data provided by ULBs under following parameter

Refining the policy based on these observations

Relevant and up-to-date techniology

User-friendly technology

API's are easily available and accessible

Data Security

Number of ULBs covered by dataset

Numbers of functions within a department covered by dataset

KMDS can appoint a Chief Data Officer for all of the below functions and to supervise the data officers of the ULB's. This officer must engage with the municipal data officers to assess how KMDS can support the data needs of the ULB's. Each urban local body should start by carrying out an assessment study of their current data with regards to the following parameters:

    a.   Is the technology relevant and up to date?

    b.   Is it user friendly?

    a.   Are the API's easily available and accessible?

    a. How does the data fare in terms of security?

    b. How many ULB's does the data covers (breadth of coverage) and how many functions or services within a department it covers? (depth of coverage)

    c.   Findings of the above should be presented to commence preparations of refining the policy

It will be their mandate to oversee the creation of a comprehensive inventory of datasets held by each ULB which is published to the central open data location and is regularly updated. Significantly, such an officer can also function as a data ombudsman for the public, fielding public feedback and ensuring the policy is included into a long-term data strategy. The development and implementation of these practices shall be overseen by the Nodal Officer, reporting to the Director, Directorate of Municipal Administration [or to Director, Directorate of Municipal Administration's designee]

<u>Open Data Observatory</u>

The Observatory can be the centre for analysis of all the data emerging from the ULB's where a team of data analysts predict trends in various aspects of urban governance. This initial trend analysis can

then be discussed in a forum with the data officers of the ULB's and KMDS in a monthly meeting, leading to a brainstorming about the improvement of various urban governance issues by utilizing such data. The Observatory can also be categorized by teams on specific aspects such as waste management, citizen safety, water management etc.

[KUIDFC to decide if the above structure is feasible and if it may be included within the policy framework]

b) <u>DATA MONETIZATION</u>

As elaborated in the earlier parts of this policy document, release of urban local bodies data in the public domain would facilitate innovation, ensure progress and improve quality of living within the city. However the data released in public domain attains value once it is processed and analysed, which will then allow researchers and organizations to study such data and propose policy trends or engage in initiatives based on such trends. However globally there has emerged a concept of allowing cities to monetise the data generated by them and use the proceeds to fund various infrastructure projects.

Monetisation of data is the ability of urban local bodies to create a value proposition for the data that is generated and make it viable for businesses/individuals to purchase. This policy document provides a structure for urban local bodies with the assistance of the Government of Karnataka to monetize some of its data, profitably, while not compromising on its objective to make municipal data open and accessible.

Data monetisation vs open data

Urban local bodies and Government of Karnataka is committed to ensuring that data in its purest form is open and accessible to every resident of the state. However to enable data to have commercial value, urban local bodies and State Governments must process and analyse the data. Urban local bodies either on demand or upon anticipation of commercial trends may connect two or more datasets, analyse and process the same. Such activity would lead to creation of value for the data enabling its monetisation.

However, data published as part of the open data objective will be singular, non-connected and left to an individual to create value from such data. Hence the extent of monetisation by an urban local body would depend on its ability to create value to the data. This policy document provides a detailed pathway for urban local bodies to create such value.

Example-

As part of the open data objective of each urban local body, they would be required to release data pertaining to roads and drainage pipes within the city. To monetize such data, urban local bodies may analyse the roads and drainage pipes data and provide internet service providers various routes for laying of optical fibre cables.

Therefore, once the data is analysed and it is made commercially viable, it provides an opportunity for urban local bodies to work with the private sector to enable better business value and improve municipal revenues.

<u>*Creating value for data*</u>

The 2018-19 Economic Survey of India recognizes that for governments across the country to create value for the data that is generated, compliance to the following characteristics would be crucial.

First, merging of datasets is an important characteristic for creating value for data. *For instance,* merging of data pertaining to solid waste with trade license would potentially create value to understand sectors which generate bulk waste and would facilitate innovation and creation of business models in the areas of waste treatment. Additionally, combining of distinct datasets can be useful in designing welfare measures in partnership with the private sector.

Second, data needs to cover a critical mass of individuals/firms so that comparisons and correlations can be assessed among individuals/firms to generate useful policy insights. *For instance*, data pertaining to solid waste generated would be valuable if it covers multiple areas and users. Therefore diversity in data would be critical to ensure the data has value.

Third, data must have a long enough time-series so that dynamic effects can be studied and employed for policy making. For instance data pertaining to solid waste generated would be valuable if it is collected across various seasons of the year.

Inter-linkage of datasets

The unique municipal framework in Karnataka has certain local functions such as transport and housing being performed by parastatal agencies. For data to be useful and commercially valuable the data from those parastatal agencies should be analysed holistically. For instance, for real estate agencies to be able to decide locations for their upcoming projects it would be imperative that data pertaining to road transport, housing and master plan are together analysed. Hence for the purposes of Monetising data it is proposed that the urban local body take up the initiative to enter into a memorandum of understanding with the parastatal agencies or enable the same through a statutory provision and share revenues from the data accordingly. Basis the memorandum of understanding the urban local bodies will process the self generated data and those collected from parastatal agencies in accordance with the open data policy.

[KUIDFC to decide on the appropriate framework on inter-linkage of datasets and if the urban local bodies should take up the lead role for coordination of data only for those types data which may be monetised.]

Purpose of urban local bodies to coordinate, analyse and process data

The purpose of urban local bodies to involve itself in analytics of data is to create an additional source of revenue and also utilize its data to facilitate economic activity within the city. This policy recognizes *'data lead financing of smart infrastructure'* as a goal for urban local bodies to reach by _____. Under this goal, all smart cities in Karnataka should be able to finance upto 5% of its

projects by monetizing data by the year _____, 10% by the year ____ and 15% by the year _____. The Government of Karnataka in addition to providing suitable technical assistance to achieve this goal may also provide incentive grants to urban local bodies for achieving this goal.

[KUIDFC to decide if the above suggestion is viable and if it may be included within the policy framework]

Regulatory structure and framework for monetization

Requirement to anonymize

As explained in the earlier parts of this policy, regulation of usage of non-personal municipal data is within the legislative competence of the State Government. However, the question emerges if the sale of municipal non-person data encounters any regulatory hurdles. Currently the Central Government does not have any legal provisions for the sale of non-personal municipal data. Under the proposed Personal Data Protection Bill, 2020 the Government recognizes the importance of anonymization of data prior to its sale.

Framework for sale and sharing of data.

Sale of non-personal data by municipal bodies currently does not attract any legal sanctions. However, it is proposed that the Karnataka Municipal Corporations Act,1976 may be amended to allow urban local bodies to specifically permit urban local bodies to monetize data and coordinate with other parastatal agencies for the collection of data.

Establishing a data analytics team at each smart city.

The Government of India's Smart City Mission requires each smart city to set up a command and control centre, which acts as the hub for undertaking all technology related functions of the smart city. To enable effective monetisation of data at each smart city/urban local body, a dedicated analytics team may be established by the respective Special Purpose Vehicle for smart cities.

DATA QUALITY

Data quality goes beyond merely the accuracy of the data, it also refers to elements of credibility of the data source. There are multiple aspects to making data ready for publication such as ensuring the appropriate context of its publication. ULB's store several kinds of data, not all of them are collected in the context that they will be published in. The methodology of collecting data, if not suitable, can give rise to a data that is not representative of the ground reality.

Minimization of errors

The first step is to minimize the errors during data collection itself. To do this, the ULB's can first remove errors or discrepancies within the primary source itself. Such as if surveys collected on citizen's welfare needs should be cross checked from them and any other source such as government data already collected to ensure that it is consistent. If not, the primary source of such data should be verified again. Secondly, they can include essential metadata for each dataset along with contextual information where available. The data presented should be labelled with descriptors that state its format, date etc as mentioned earlier in the policy. Further, data should be auditable in order to improve transparency and authenticity in data collection and publishing process, put checks and balances to remove any scope of manipulations. Lastly, system generated data should be encouraged which captures data electronically through process mediated mechanisms or hardware/sensors. Especially when seen in consonance with a Smart City mission, there will we be several points of data being captured by devices which can be utilized as a source of open data.

Data Validation

Data can be then validated in the following ways:

a) By corroborating with various ULB departments – Using cross-linking and correlation to ensure common attributes of the open data are consistent
b) Aggregating data at the zone/ward level – Ensuring data is available at granular level that adds up to the total value being projected at the higher institution of governance
c) As per benchmark data reported on other public data portals or other reputable sources.

Data Updation

Public data sets should be updated as often as is necessary to preserve the integrity and usefulness of the data sets to the extent that the agency regularly maintains or updates the public data set. Each ULB, in collaboration with KUIDFC, shall determine the frequency for updates to a dataset, and the mechanism to be utilized. To the extent possible, datasets shall be updated through an automated process to limit the additional burden on agency resources.

RE-USING DATA

This section describes the mechanisms through which it is ensured that the open data published will be available for different stakeholders to consume, modify and further utilize. The primary purpose of this policy is to ensure the continuity of the data being released which is ensured only when the public is aware of the conditions of its re-usage, the conditions should reassure and encourage them to engage with the data. It should clarify any doubts they could potentially have on using the data for their purpose and benefit. Hence, the ULB's state that:

a) Datasets shall be made available to the public on an open license basis. An open license on a dataset signifies there are no restrictions on copying, publishing, further distributing, modifying or using the data for a non-commercial or commercial purpose.

b) Each ULB will assure that data supplied by third parties (developers, contractors, consultants) are unlicensed, in a prevailing open standard format, and not copyrighted except if otherwise prevented by legal considerations.

c) The purpose of this waiver is to ensure that users understand their rights and are thereby encouraged to use this community asset in new and valuable ways. These include the abilities to retrieve, download, index, sort, search, and reuse data for novel purposes.

d) It is possible that the Open Data may contain errors, inaccuracies or typographical errors, which may be subsequently corrected. In no event shall the ULB or its officers, employees, or agents be held liable for any direct, indirect, incidental, consequential, punitive, special, direct or other damages whatsoever arising out of or relating to the Open Data or the user's use of the same, including all claims and claims by third parties, for loss or damages arising from an error, inaccuracy, or other defect in the Open Data, or information contained in the open data platform.

e) It may require a third party providing to the public any public data set, or application utilizing such data set, to explicitly identify the source and version of the public data set, and a description of any modifications made to such public data set. Such data citation can be specified, where needed.

[KUIDFC to decide which mode of attribution to include within the policy]

SECURITY AND PRIVACY OF DATA

Data classification

Earlier in the policy, we have specified a mode for access to data but this section deals with the classification of data as it comes in. The question of access is designed to be based on this foundation of a classification. The Central Policy, NDSAP does not prescribe a data classification system, leaving it up to the states to do so. Hence, we recommend a system that can be used for any data that is generated or gathered by the urban local bodies to be graded in the following manner:

1. Public – Data available for public consumption

2. Internal Use – Information which can only be disclosed to ULB employees for managing operations or delivery of public services on day to day basis

3. Sensitive – Data regulated by any Municipal;/State/Central law or regulation like privacy

4. Protected – Data which needs to be protected for eg: identity of citizens and disclosure/notification needs to be issued by ULB in case of any breach or loss of data

5. Restricted – Data which could lead to threat to life or loss or public assets or critical infrastructure

Privacy

- Data licensed to the city by another person or entity shall not be made public under this chapter unless the person or entity licensing the data agrees to the public disclosure

- Proprietary, privileged, and other information protected from disclosure by law, ethical standard or contract shall not be disclosed.

- Such disclosure shall be consistent with applicable law, related to security, privacy and confidentiality and no personally identifiable information shall be posted online unless the identified individual has consented to the posting or the posting is necessary to fulfil the lawful purposes or duties of the agency.

- Data anonymization should be carried out wherever applicable to protect sources and objects of data where needed. Personal Identifiable Information cannot be published by ULB's on platforms under any data sets. Data sets must be anonymized before publishing which means it should be presented in such a manner that the data subject is not or no longer identifiable. In other word, anonymization means excluding any personal identifiers from data sets.

- The Data Protection Bill empowers the Data Protection Authority (DPA) to establish standards for anonymization and consequently, as and when that happens, the process followed by the ULB's will need to be in consonance with the DPA's recommendations.

- However, to begin with, the suggestions from the United Kingdom(UK) Data Service can be seen as guiding steps, involving the following concepts of:

    ● **Direct identifiers** such as names, postcode information or pictures

    ● **Indirect identifiers** which, when linked with other available information, could identify someone, for example information on workplace, occupation, salary or age

They suggest that quantitative or qualitative data can be anonymized by:

1.Find and highlight direct identifiers

    ● Quantitative: visually scan variables

    ● Qualitative: read the transcript

2.Assess indirect identifiers

    ● Can the identity of a participant be known from information in the data file

    ● Can a third party be disclosed or harmed from information in the data file

3.Assess the wider picture

    ● Quantitative: run descriptive statistics and crosstabs to find unique cases and combinations of variables that can identify an individual in the dataset

    ● Qualitative: which identifying information about an individual participant can be noted from all the data and documentation available to a user

CITIZEN ENGAGEMENT

Initiatives from the ULB's

Merely publishing data is not enough for ULB's to foster a data culture, to do so they will have to involve stakeholders though different channels. Forming alliances in the open data community through activities such as conducting hackathons on the city's open data, sharing from related data sets from stakeholders such as civil society organizations in order to facilitate better engagement, seeking data submissions from individuals and entities on topical matters such as in today's current scenario, information on pandemics and municipalities. Data Challenges for students and those in academia have also proved to be a popular method of involving citizens in open data efforts.

How can citizens engage on the data platform?

    A) The open data platform can incorporate an on-line forum to solicit feedback from the public and to encourage public discussion on open data policies and public data set availability on the web portal.

B) It shall include a mechanism for the public to give feedback on and assess the quality of published information, provide input about what information should be a priority for publication, and provide input on the open data policy itself.

C) On a quarterly basis, the ULB's will review the use of the open Data Policy through comments from staff and the public, including from individuals and firms who have successfully developed applications using data from the platform.

D) IT professionals across departments will collaborate and take proactive efforts to publicize open datasets, establish useful applications employing these datasets, and collaborate with local entities (civic-minded developers, universities, corporations, start-ups, non-profits, civic organizations etc.) to experiment with new technologies, visualizations, and applications with the goal of identifying cost-effective solutions.

Analysing Feedback by citizens

Within one year of this policy, the Chief Data Officer shall publish an annual Open Data Report. The report shall include an assessment of progress towards achievement of the goals of the Urban local bodies 's Open Data Program, an assessment of how the Urban local bodies 's open data work has furthered or will further the Urban local bodies 's programmatic priorities, and a description and publication timeline for datasets envisioned to be published by the Urban local bodies in the following year. During the review and reporting period, the Chief Data Officer should also make suggestions for improving the Urban local bodies 's open data management processes in order to ensure that the Urban local bodies continues to move towards the achievement of the policy's goals.

ANNEXURE 1 – QUESTIONS TO GAUGE ULB's DATA NEEDS

1. Does the ULB publish any data on its website? If yes, what kinds of data and how often is it updated?
2. Does the ULB currently have an open data strategy or policy in place?
3. Are you currently in the middle of developing such a policy, if not in place already?
4. Do you have any other policies that cover open data such as transparency policy or government data disclosure policy?
5. Would you like training in how to formulate and implement a state- wide open data policy or is having a policy sufficient?
6. Would opening of data be beneficial to the ULB?
7. If the ULB's data is periodically and proactively disclosed, would the residents of the ULB benefit from it or utilize it?
8. Does the ULB forecast any commercial demands for any of its data?

9. Does the ULB currently have any policies or initiatives in which you're engaging in public participation with citizens?
10. Does the ULB currently have a data officer?
11. To improve data disclosure and maintain an active and user friendly website what support would the ULB require from the State Government?
12. What technical challenges do you see in implementing an active open data framework in your ULB?