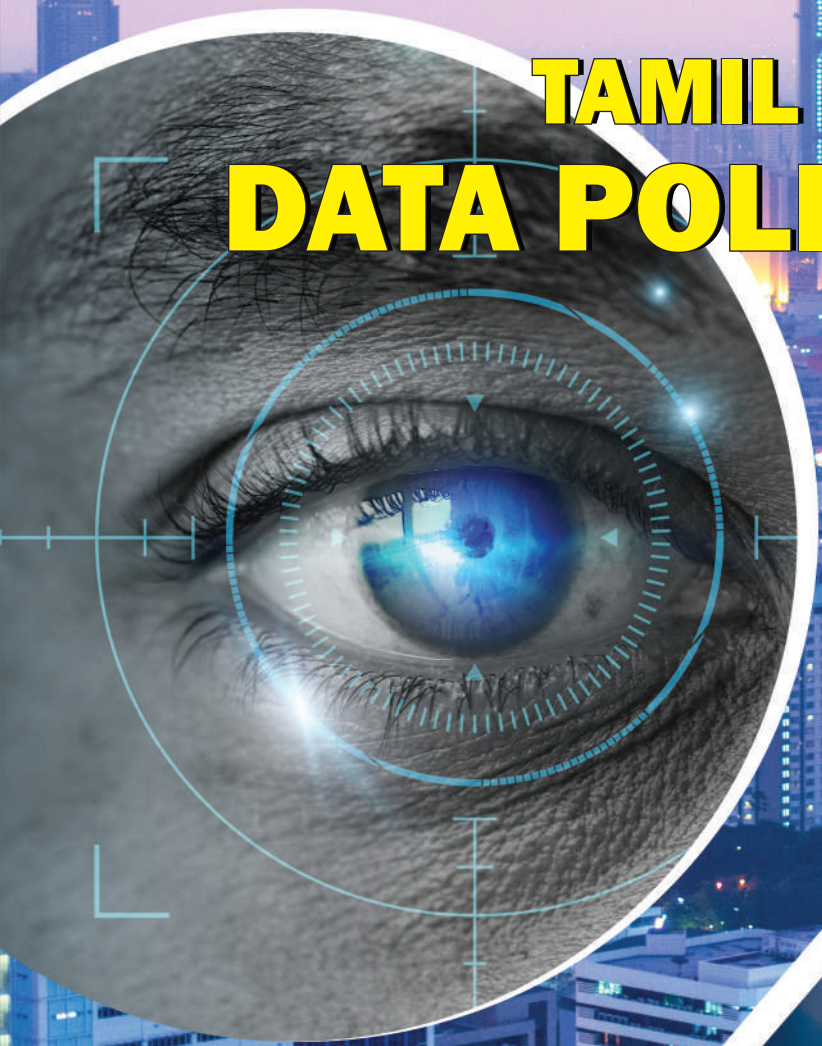
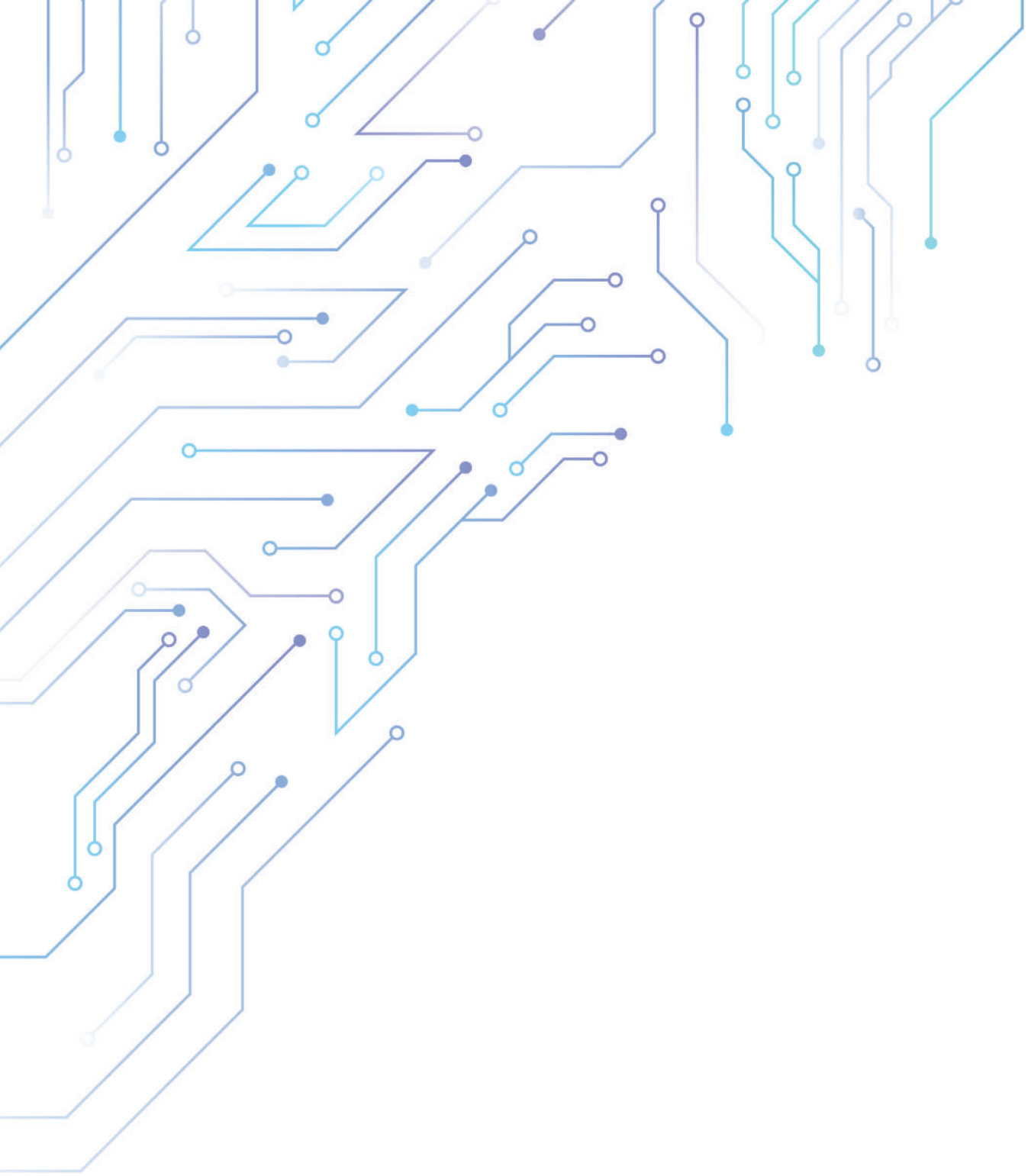
Information Technology Department

TAMIL NADU
DATA POLICY - 2022



TNeGA

Government of Tamil Nadu



TAMIL NADU
DATA POLICY-2022

INFORMATION TECHNOLOGY DEPARTMENT
GOVERNMENT OF TAMIL NADU

Message

Data is the new oil. Tamil Nadu Government is acutely aware of the need for good quality data in enabling Good Governance to the citizens. The Economic Advisory Council has encouraged us to use data for improving governance. The use and sharing of open data will also improve transparency. We have committed ourselves to use data to improve decision making, transparency and providing services to the citizens in a pro-active manner. Information Technology Department is the nodal department to collect, process and share data. I am happy that 'Tamil Nadu Data Policy' that has been prepared by it will enable various departments to use data for making better decisions and improves transparency.

Thiru. M.K.Stalin

Hon'ble Chief Minister
Government of Tamil Nadu

Message

Tamil Nadu Government has been proactively encouraging the use of data for improving the lives of citizens and ensure data-driven decision making. This goal can only be met with good quality data captured by various departments as they roll out their programs and by digitizing their existing data collection. There is also significant scope for use of open data for enhancing transparency and private sector to provide value added services. The ‘Tamil Nadu Data Policy’ is a comprehensive framework that covers all these aspects and I am sure it will provide an excellent base for effective governance.

Thiru. Mano Thangaraj

Hon’ble Minister
Information Technology
Government of Tamil Nadu

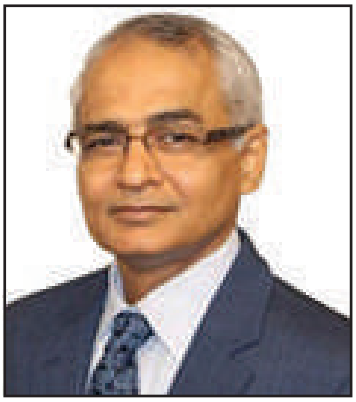
Message

Tamil Nadu has been a State that has given birth to many pioneering schemes of public importance that have been replicated even nationally. These can be attributed to astute leadership and ground level connection of the political leadership and a responsive bureaucracy. However, the times are changing rapidly. Rapid technological changes are driving even more rapid change in our social lives and expectations. Therefore, Governance must respond with equal rapidity. Further, Black swan events such as Covid pandemic have exposed the fault lines in our public systems and stressed our fiscal capacity very adversely.

Therefore, the need to rely on data and design policies that can respond to rapid changes and simultaneously meet the expectations of the citizens has become even more acute. This requires sharing of data at the State level to improve cross learning and effective design of schemes not only across departments, but also with NGOs and research agencies, while preserving the privacy of citizens. I am happy to see that 'Tamil Nadu Data Policy' will meet these data needs of data driven decision making and transparency.

Dr. V. Irai Anbu IAS

Chief Secretary
Government of Tamil Nadu



Foreword

The Government during the process of governance and in delivery of services to citizens, businesses, internal stakeholders, elicits, collects, processes, and stores vast amount of data. This data can aid policy makers in crafting evidence-based, data-driven policy formulation and strategic decisions.

Tamil Nadu Data Policy is an outcome focused policy that aims to improve policy making and benefits disbursement through innovative use of data to improve scheme implementation, encourage value added services by data sharing, improve access and quality of services, encourage creativity and innovation.

Tamil Nadu Data Policy is based on thirteen guiding principles. It is not only a “data policy” but also a framework for “data for policy”. It accordingly not only provides for data classification but allows for sharing of data for public good, is sensitive to individual privacy concerns, and allows for a grievance mechanism.

Data management is an evolving field with new tools and technologies being developed constantly. In such a dynamic environment, to ensure equity and ethical use of data, and retaining its relevance, the Tamil Nadu Data Policy (TNDP) is envisioned as a living, dynamic policy that would proactively adapt to changes in society and technological advancements through citizen feedback and industry experts on the subject to fulfill its overarching goal namely, “data for public good.”

Dr. Neeraj Mittal IAS
Principal Secretary (IT)
Government of Tamil Nadu

TABLE OF CONTENTS

ABSTRACT

Information Technology Department – Harnessing Data for effective decision-making in the State - Tamil Nadu Data Policy, 2022 – Approved - Orders - Issued.

INFORMATION TECHNOLOGY (E1) DEPARTMENT

G.O.(Ms). No. 16 **Dated : 10.03.2022**
பிலவ, மாசி – 26
திருவள்ளுவர் ஆண்டு–2053

Read:

1. From the Director of e-Governance & Chief Executive Officer, Tamil Nadu e-Governance Agency, Letter No.15/61/2020-CoE-CeG/TNeGA, dated 06.12.2021.
2. Government Letter No.1746584/IT(E1)/2021, dated 13.12.2021.
3. From the Director of e-Governance & Chief Executive Officer, Tamil Nadu e-Governance Agency, Letter No.15/19/2021-CoE-CeG/TNeGA, dated 27.12.2021.

****

ORDER:

Government of Tamil Nadu is committed to the vision of evidence-based policy making to improve governance, transparency and providing services to the citizens in a pro-active manner. In the process of delivering services to its citizens and businesses, the Government collects, processes, stores and analyzes large amount of data. This data, can be an effective aid in data-driven decision-making.

2. The Director of e-Governance & Chief Executive Officer (DeG & CEO), Tamil Nadu e-Governance Agency (TNeGA) in the letter 1st read above, has prepared and sent a draft Tamil Nadu Data Policy (TNDP) to improve policy making and disbursement of benefits through innovative use of data.

3. In the Government letter 2rd read above, the draft Tamil Nadu Data Policy (TNDP) has been communicated to the major stakeholder Department, i.e. Planning, Development & Special Initiatives Department for acceptance. Further, in order to obtain view of relevant stakeholders, the DeG & CEO, TNeGA has been directed to consult with various stakeholders, i.e. Department of Economics and Statistics (DES) and Planning, Development & Special Initiatives Department and then send a revised policy by incorporating the comments.

4. In the letter 3rd read above, the DeG & CEO, TNeGA has stated that a consultation meeting was conducted on 20.12.2021 with various stakeholders from multiple agencies like Government Departments (Planning, Development and Special Initiatives Department and Health Department), Academic Institutions (Anna University and Chennai Mathematical Institute), Central Government Agency (NIC), NGO's (J-PAL and Indus Action) and Industry Partner (Convergence Catalyst). Based on the discussion,

the DeG & CEO, TNeGA has sent a revised version of Tamil Nadu Data Policy (TNDP) to Government by incorporating the suggestions.

5. Accordingly, the Government, after careful consideration, has formulated the Tamil Nadu Data Policy (TNDP), 2022 to support policy making, improve scheme implementation, encourage value added services, improve access to and quality of services. This Policy provides for data classification, sharing of data for public good, taking into account privacy concerns, and allows for a grievance mechanism. It will help in setting standards of sensitive data, data exchange, and rules governing access and use.

6. Tamil Nadu Data Policy, 2022 has been brought about to utilize the large amount of data that is generated in the process of delivering services to citizens and businesses. The Government of Tamil Nadu has already taken several steps to use the data:

- Data Analytics Unit,
- Data Purity initiative by TNeGA,
- Centre of Excellence in Emerging Technologies (CEET) at TNeGA,
- Tamil Nadu Open Data Portal (https://tn.data.gov.in/),
- An inter-department data exchange platform - the State Family Database (SFDB),
- Data eco-system for Data Driven Decision Support System (DeTN),
- Spatial Data Infrastructure through Tamil Nadu Geographical Information System (TNGIS),
- Population Health Registry by Health Department.

The soul of **Tamil Nadu Data Policy (TNDP)** is to use of "**data for public good**".

Objectives of TNDP:

- Develop data capabilities to address all phases of governance, namely - policy making, administration, scheme review, policy redesign, aimed at developing a data-driven decision-making culture.
- Promote the data analytics capability of all the departments.
- Minimize exclusion/inclusion errors in Government schemes.
- Drive data based assessment of efficacy of schemes.
- Making data publicly available for policy research and for improving transparency.

Benefits of TNDP:

- Maximise use of data
- Minimise time/cost of data collection
- Integration
- Better decision-making and inter-department co-ordination
- Equitable access
- Fostering Innovation
- Job creation
- Minimise errors in beneficiary selection
- Protection of citizen privacy rights
- Transparency

..3

Implementation of TNDP:

- A **State Level Empowered Data Governance Committee (EDGC)** headed by the Chief Secretary will provide strategic guidance for the State data policy framework.
- A **Data Inter-Departmental Committee (DIDC)** headed by the CeO, TNeGA, the State's Chief Data Officer (CDO) will take operational level decisions emanating from the Policy.
- The **Department Data Officer** in each department shall be responsible for its department data collection, decisions relating to open data, storage, compliance to the Policy and carry out the decisions of the Data Inter-Departmental Committee (DIDC).
- The Tamil Nadu Data Policy is a dynamic document that would need to adapt to changes in society and technological advancements in data creation, management, sharing and associated challenges. The Information Technology Department, Government of Tamil Nadu will continue evolving the Policy further from time-to-time in tune with the technological advancements and State's requirements.

7. The Tamil Nadu Data Policy, 2022 has been issued with the concurrence of multiple stakeholder departments, including Planning, Development & Special Initiatives Department and Department of Economics and Statistics. The Policy document is annexed with this Order, for compliance by all departments.

(By Order of the Governor)

NEERAJ MITTAL
PRINCIPAL SECRETARY TO GOVERNMENT

To

All Departments of Secretariat, Chennai - 600 009.
The Additional Chief Secretary to Government,
Planning, Development & Special Initiatives Department,
Secretariat, Chennai – 600 009.
The Director of e-Governance / Chief Executive Officer,
Directorate of e-Governance / Tamil Nadu e-Governance Agency,
2nd & 7th Floor, P.T.LEE Chengalvaraya Naicker Building,
Anna Salai, Chennai - 600 002.
The Commissioner,
Department of Economics and Statistics,
DMS Complex, 259, Block –II, Teynampet, Chennai -600 006.

Copy to:

The Special Personal Assistant to Hon'ble Chief Minister,
Chief Minister's Office, Secretariat, Chennai – 600 009.
The Special Personal Assistant to Hon'ble Minister for Information Technology,
Secretariat, Chennai – 600 009.

..4

The Personal Assistant to Hon'ble Minister for Finance and Human Resource Management, Secretariat, Chennai –600 009.
The Principal Private Secretary to Chief Secretary to Government, Secretariat, Chennai-600 009.
The Private Secretary to Principal Secretary to Government, Information Technology Department, Secretariat, Chennai-600 009.
The Information Technology (OP) Department, Secretariat, Chennai-600 009.
The Content Creator/Moderator/Nodal Officer, Information Technology Department, Secretariat, Chennai-600 009.
SF/Sc

// Forwarded / By Order //

10/03/2022

Section Officer.

10/3/22

1 PREAMBLE

We live in an age where cost of computing, cost of data generation and cost of storage are continually decreasing, resulting in a proliferation of data in practically every sector. Be it banking, manufacturing, real estate, travel, health or any other sector, the impact due to data-based business models and digital disruption has been huge over the last 10 years.

This trend will accelerate as we move forward, and governance should utilize this change as an opportunity to impart reforms that are user friendly from citizen, business, and government standpoint. Traditional paper-based governance mechanisms have been replaced by digital systems in most e-Governance systems. However, the true potential of data in Government sector is yet to be realized fully. This is because of multiple reasons as follows:

(i) Access to systems and data in government remained confined to individual departments in a largely disconnected manner. The respective government departments are yet to come up with mechanisms to quantify the data from reporting and decision-making standpoint, to have periodic audits (internally / assessment by a professional third party), to classify useful (such as data in their production environment) and redundant data (data to be archived / rejected from department data repository).

(ii) Lack of common guidelines on what data can be shared, how to make data available for use within and across departments and can be made available for external parties.

(iii) Limited use of data for impact assessment, monitoring, and mid-course corrections.

(iv) Absence of data-based governance that prevents evolution of a data-driven culture throughout.

The soul of Tamil Nadu Data Policy (TNDP) is to enable use of **"data for public good."**

Following NDSAP (National Data Sharing and Accessibility Policy) 2012, Government of India came up with Open Government Data (OGD) Platform to promote publishing of Government owned and publicly shareable data in machine readable form. Government of Tamil Nadu have taken up OGD initiative by publishing several datasets, but this effort has been hampered due to lack of a structured mechanism to update and publish such data sets.

The Government of Tamil Nadu have committed to evidence-based approach to policy making.Though the data collected by different departments have made its way through OGD initiative to external stakeholders to some extent, the original data which is richer and has beneficiary level information has not been used much by the Government. There is tremendous scope for use of this richer data for enabling data-driven decision and policy making. This data hasn't been useable due to various reasons. Firstly, data has resided in silos in individual departments. Secondly, departments have been collecting data in a non-standardized manner.Thirdly, the departments vary in the extent of e-governance and hence its data may or may not be digital. This data policy solves the first challenge by providing a framework of sharing and use of sensitive data for public good.

For enabling use of "data as a public good", the entire administration of Government schemes will have to be looked through a '**data lens**'. When we use the term data in an unqualified manner in this Policy, it covers both open data and data which is not open.

The data can be static or dynamic as is the case when it is transactional data. The data should be sufficient and of adequate quality to address requirements from a policy design perspective, scheme formulation, scheme implementation, assessment of scheme effectiveness and policy redesign.

Data becomes useful from a Policy design and implementation and effectiveness perspective only when it can be combined with insights from data analysis, comparison and validation with independent surveys and related data sources, and evidence from research. The Tamil Nadu Economic Advisory Council has highlighted the need for additional datasets through surveys and administrative data combined with research as fundamental to economic policy making and decisions throughout the policy making lifecycle (design – execution – course correction and measurement of outcomes). To this end, it is envisaged that regular and live data collected may be made available (in a secure manner, without compromising privacy and security) to both researchers and members of the public for research and non-commercial use, to inform policy making and improve effectiveness of government programs with adequate safeguards.

This Policy provides the framework and guidelines for publishing of open data and collation of data, secure access to non-open data to catalyse the research and policy making by line departments, the Data Analytics Unit, and external, bonafide research and statistical agencies collaborating with the government to strengthen policy design and implementation.

The roles and the use of data at each step of Policy design are captured in the table below:

Policy Design Phase *(By Line Dept. / PD&SI Dept.)*	**Steps** *(By Line Dept. / PD&SI Dept. directly and through collaborations with external agencies)*	**Data Support** *(By TNeGA)*
Policy Formulation	Target beneficiary group identification and quantification. Quantification and sizing of target group across multiple dimensions. Cross department analysis of target group with respect to other schemes. Reach-out strategies for various target groups based on different demographic parameters.	Share the available data in the repositories in a secure manner ensuring data privacy and protection of sensitive personal data. Further, such sharing of data will be shared as public good for policy formulation and as specified in the Policy.

Scheme Rollout and Administration	Developing Rollout Strategy - Developing right rollout and PR Strategy for schemes can be done with right data analysis.	
	Efficacy Assessment - A periodic sample survey-based assessment could be done to determine the effectiveness of schemes.	Data layer to support these type of sample surveys. Help in conducting Phone / electronic surveys.
	Corrective Measures Identification - Different scheme implementations would need course corrections in between depending on efficacy.	Data analysis and necessary improvements to underlying data infrastructure, visualizations etc. There will be an emphasis on data quality and accuracy.
	Leakage assessment	Inclusion and exclusion errors detection based on data.
Scheme Analysis, Evaluation and Policy Redesign	Impact assessment – Concurrent and at the end of a scheme, an analysis should be done to quantify the total impact and effectiveness.	Data layer to support these type of sample surveys. Help in conducting Phone / electronic surveys.
	Root cause assessment - If certain schemes do not result in intended outcomes, then a root cause analysis should be done with the right data sources to identify and fix gaps. Similarly, if a certain scheme worked well, the key success factors should be identified with the right data analysis.	Data Analysis
	Redesign policy - Redesign of existing policy could be done based on outcomes of impact assessment and root cause assessment as well as other external data sources.	Data Analysis

Tamil Nadu e-Governance Agency (TNeGA) will function as a repository of all data, perform data analytics, and help departments / DAU in policy making and analysis.

This requires a quantum leap in use and quality of Government data in all aspects – citizens data, master data, transactional data that is generated through various e-Governance

systems. The TNDP brings in a framework through which these could be achieved so that data and actionable insights become the cornerstone of policymaking. The policy also covers principles and roles of various stakeholders in publishing open data.

Governments are the largest generator of data across multiple dimensions. As this data is generated by public funds, it is essential that this is used for maximum benefit of public. The National Data Sharing and Accessibility Policy (NDSAP), 2012 highlights that evidence-based planning of socio-economic development processes require quality data, and the need for a policy to leverage data residing among the entities of the government, by facilitating sharing and utilization of data generated by these entities. The open data could be leveraged by private sector to develop value added services as well. The Policy intends to achieve this in a non-discriminatory manner while safeguarding the privacy of individuals and institutions.

The Hon'ble Finance Minister in his budget speech for 2021-22 has laid emphasis on Data Centric Governance reforms and flagged lack of appropriate beneficiary data as a limiting factor in efficient scheme administration and delivery. Budget speech also has laid emphasis on cross-department data analysis as a focus area. The Policy develops these focus areas.

In 2016, the Government of Tamil Nadu set up a **Data Analytics Unit** in collaboration between the Department of Economics and Statistics and the Tamil Nadu e-Governance Agency to analyze, in real time, the large volume of data available with the government to enable improved policy making for better service delivery. In order to realize the mission and potential of this initiative which attempts to provide solutions to most of the decision making problems, there is need for a State-level policy that provides a framework for all departments and data owning entities under the Tamil Nadu government, to share open data and non-open data securely, and in formats amenable for analysis and use for decision making and such that it conforms to the provisions of protection of personal and sensitive data under the Information Technology Act,2000 (Central Act 21 of 2000), the Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act, 2016(Central Act 18 of 2016) and the Rules framed thereunder. With such a framework and policy in place, the Tamil Nadu government aims to strengthen its mandate and efforts in institutionalizing evidence-based approach to decision making. *"Governments are, or could be, the most important source of open data (…) yet sustained, impactful, scaled up examples of big data and open data in developing countries are still relatively rare."* (World Development Report, 2016, Digital Dividends).

1.1 Guiding Principles

In order that the TNDP can deliver on the benefits of data sharing and accessibility to the citizens, enable departments to drive better service delivery framework through use of meaningful data sets, the Policy is based on certain core principles. These principles include:

1.1.1. Openness

An open data set is defined as follows: “A data set is said to be open if any-one is free to use, reuse and redistribute it. Open data is machine-readable, and it is also easily accessible.” Openness provides the foundation for data sharing without compromising on its purpose and utility. To the extent possible, all data should be open.

1.1.2. Privacy

Privacy is the ability to seclude information about individuals / groups such that the data does not reveal identifying information about the individual or group. The Government as the custodian of information of citizens is responsible to safeguard it. Data should be shared in a manner that preserves the privacy of citizens. The Personal Data Protection Bill, 2019 (PDP) (which is yet to become a law) lays down a few key principles such as consent and autonomy of citizens, collection and purpose limitation, security of data, and responsibility of data controllers and data fiduciaries. TNDP is written with these principles in mind to facilitate adaptability to PDP in near future, while appreciating the fact that the bill is yet to become a law. These foundational aspects will make the compliance to PDP law seamless as and when it gets approval.

1.1.3. Ethics and Equity

The Policy shall conform to the State's ethical standards. In addition, the Policy ensures equity by ensuring that data is easily accessible to all irrespective of means or social status. "Non-discrimination" refers to who can access data and how they must do so. Barriers to use of data can include prior registration, membership requirements or unreasonable waiting period for accessing data after it is requested. Sometimes, organizations may resort to the use of "walled garden." This means that only certain applications are allowed access to data. By avoiding all these inequitable means, the State creates a level-playing field among the citizens and NGOs, encouraging innovation and creativity, all the while respecting citizen privacy.

1.1.4. Flexibility

The Policy shall promote flexibility which is the ability to ensure openness despite changing nature of the data – variables, fields, size, formats, and other characteristics of the data.

1.1.5. Transparency

The Policy will ensure transparency by providing free and easy public access to open government data with clear traceability to sources of data and information about any intermediate data transformations, if any. By providing transparency, the Government allows the usefulness and utilization of the data for public good to be maximized as mandated by TNDP.

1.1.6. Legal Conformity

The Policy will conform to all the laws of the land including the Central and State laws on privacy, data and information protection. This will remove all legal barriers to the use of data by citizens for public good.

1.1.7. Protection of Intellectual Property

One of the objectives of the Policy is to promote innovation and creativity. Hence, it is necessary to respect the intellectual property rights of citizens and the owners of data by restricting access to IPR protected data. This calls for the introduction of checks and balances to ensure that data is protected by IPRs is not hosted as Open Data.

1.1.8. Interoperability and Standards

When systems are designed, it should be built with a perspective of interoperability from ground up. This results in more efficient data flow between systems. Data consumers should be able to use the data easily, without expending unreasonable amounts of time and energy on data access, cleansing and processing to convert it into usable form.

1.1.9. Quality

The usefulness of data is primarily determined by its quality. High quality data is free from anomalies such as missing data, wrong attribute values, junk characters and the like. Such anomalies would inhibit further processing without significant investment on data cleaning and transformation. The policy defines guidelines to ensure that the shared data meets minimum quality requirements.

1.1.10. Security

In order that the same data is available to all parties that seek to access it, it must be immutable, encrypted and de-identified, wherever possible. This calls for very high-level security during hosting and dissemination to prevent data manipulation and adversarial attacks or misuse or unauthorized access to data.

1.1.11. Accountability and Formal Responsibility

The Policy provides for guidelines that ensure accountability and responsibility with respect to the contents and sharing of open data.

1.1.12. Sustainability

The goal of sustainability is to ensure that open data sharing is not hindered due to external factors. This implies sustainability on all fronts that include technical, economic, financial, legal and other relevant criteria.

1.1.13. Usability

The data that is being collected by various departments must be in meaningful and usable format in line with data policy guidelines, principles of data sharing, classification and sharing. This would in-turn will determine the focus and efforts on generation and upkeep of the relevant data.

2 NEED FOR THE POLICY

In the process of delivering services to its citizens and businesses, the Government collects, processes, stores and analyzes large amount of data. This data, with appropriate usage will be able to aid policy makers in data-driven decision-making. Further, by allowing access to open as well as sensitive but anonymized data will create a level-playing field and allow firms to create/enhance service delivery platforms for public good. There are gaps in executing data centric governance:

- Planning, policy making and service delivery are not data driven.
- Lack of availability of data for public-private partnerships for innovation in governance.
- Lack of availability of high-quality data for cutting-edge research and academic excellence.

Bridging these gaps require availability of high-quality digitized data. Although such data is available within various Government Departments and Agencies, its use has been constrained by the absence of clear guidelines governing sharing and use.

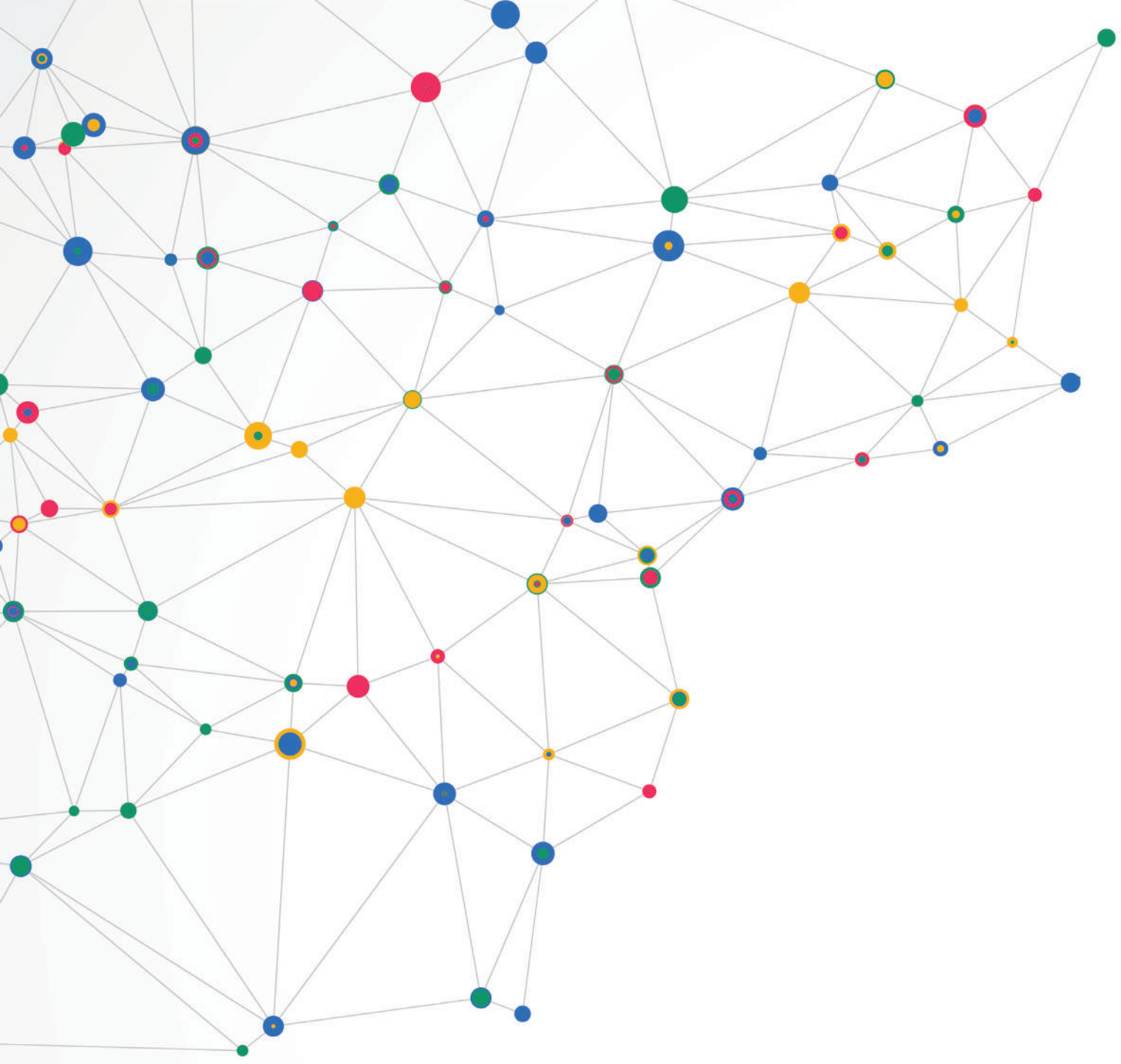
This Policy provides a comprehensive framework for data owners within the Government of Tamil Nadu. This Policy will provide guidelines to make granular and machine-readable data accessible in a transparent manner to all stakeholders while ensuring privacy and in compliance with existing rules and laws.

To this end, the Government of Tamil Nadu has already taken pro-active initiatives as detailed below:

- In the year 2016, the Tamil Nadu Government set up a 'Data Analytics Unit.'
- Data purity initiative was launched by which all the data with Government departments is now to be shared with TNeGA on a regular basis as ordered in G.O (Ms.) No.17, Information Technology (E1) Department, dated 23.09.2021. This mandates that TNeGA shall have a copy of all data from various departments, it shall process the data and is authorized to issue mandatory directions to all departments especially about providing copy of, corrections to, collection format/ fields, metadata, transfer, use storage and security relating to their data.
- TNeGA's Centre of Excellence in Emerging Technologies (CEET) that undertakes projects for sharing, harmonizing and exchange of data within Government departments and agencies.

- Launch of the Tamil Nadu Open Data Portal **(https://tn.data.gov.in/)**.
- Collaborations with research and academic institutions.
- Creating an inter-department data exchange platform, including creation of the State Family Database (SFDB).
- Creation of data eco-system for data driven decision support system (DeTN).
- Creation of Spatial Data Infrastructure through Tamil Nadu Geographical Information System (TNGIS).
- Creation of Population Health Registry as ordered in G.O.(Ms.) No.553, Health & Family Welfare (EAP II-2) Department, dated 26.11.2021.

This Policy will support all such data initiatives of the government and create a common framework for government databases; thus, avoiding duplication of efforts and integrate disparate datasets residing in silos. It will also set standards for classification of sensitive data, data exchange, and rules governing access and use. Policy will also help connecting and integrating State Government departments that may have common set of beneficiaries for various social welfare schemes. Departments with common/overlapping data sets will be able to have their databases synchronized from time to time, avoid duplication and identify the right set of beneficiaries.



3 DEFINITIONS

3.1 **Data** - means a representation of information, numerical compilations and observations, documents, facts, maps, images, charts, tables, reports and figures, concepts in digital and/or analog form. It covers all aspects of government functioning including G2G, G2B, G2C.

3.2 **Open Data** - is the data that is freely available to everyone to use and republish, without restrictions from copyright, patents or other mechanisms of control. Users have a worldwide, royalty-free, non-exclusive license to use, adapt, publish (either in original, or in adapted and/or derivative forms), translate, display, add value, and create derivative works (including products and services), for all lawful commercial and non-commercial purposes, and for the duration of existence of such rights over the data or information. Where Open Data is intended, it is specified in the Policy.

3.3 **Data Subject** – is a natural person who is subject of the handled data by which that person can be identified.

3.4 **Sensitive personal data** – means such personal information which consists of information about data subject relating to -

- Password
- Sex Life and Sexual Orientation
- Biometric information
- Information received by body corporate for processing, stored or processed under lawful contract or otherwise
- Financial Data, including information related to financial information such as Bank account/credit card/debit card/payment instrument details of the users
- Health Data, including physiological and mental health condition and medical records and history
- Official identifier
- Genetic Data
- Transgender status/Intersex status
- Caste or tribe
- Religious or political belief or affiliation

3.5 **Personally Identifiable Information** - is about or relating to a natural person/data subject who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute, or any other feature of the identity of such natural person, whether online or offline, or any combination of such features with any other information, and shall include any inference drawn from such data for the purpose of profiling. This includes fields such as names, addresses, phone numbers, and latitude-longitude geographic data of households.

3.6 **Data Archive** – is a place (data center, physical archive, data cloud etc.) where machine-readable data are acquired, stored, manipulated, documented prior to a cut-off past date and distributed to others for further analysis and consumption.

3.7 **Data Generation** - Initial generation / collection of data or subsequent addition of data to the same specification. This may be data specifically collected for a particular objective or may be a consequence of the different processes of the Government.

3.8 **Data set** – is a named collection of logically related features, attributes or variables including processed data or information.

3.9 **Geospatial Data** – is all data which is geographically referenced such that the location of the referenced object can be identified.

3.10 **Information** – is data embellished with a context, in other words, “Processed data.”

3.11 **Metadata** – is “data about data.” In other words, the information that describes the data source and the time, place, and conditions under which the data were created. Metadata informs the user of who, when, what, where, why, and how data were generated. Metadata allows the data to be traced to a known origin and know quality. Metadata also includes details such as the license type and data dictionary defining and explaining various parameters. Meta data consists of structural aspects such as defining the data and datasets, administrative aspects such as the processing and audit trail information, descriptive aspects such as describing the data content, and reference which is content, methodology and quality of data.

Metadata includes the description of time series and statistical data features (i.e. data source and the time, place and conditions under which the data was created) as well as the methods, procedures, concepts, variables, classification and nomenclature used, including publication date and data coverage. It shall inform about who, when, what, where, why and how data was generated. This allows the data to be traced to a known origin and known quality.

3.12 **Negative list** – is non-sharable data as declared by the departments/organizations within the Government of Tamil Nadu. This includes data/information that is expressly prohibited from disclosure as per exemptions defined under sections 8 and 9 of the Right to Information Act, 2005(Central Act 22 of 2005), the Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act, 2016, Collection of Statistics Act, 2008(Central Act 7 of 2009) and other relevant legislations.

3.13 **Data Standards** – is any application/framework that defines and embeds data handling functions (e.g. data collection, management, transfer, integration, publication) and operates on data in a manner that complies with data format and data syntax specifications produced and maintained by open standards bodies.

3.14 **Open Access** - is a principle that access to data generated from public funding should be easy, timely, user friendly and web based without any process of registration/ authorization.

3.15 **Permissioned Access** – refers to access to data for which prior permission is needed from the government in accordance with associated terms of use.

3.16 **Anonymization** - in relation to personal data, refers to such irreversible process of transforming or converting personal data to a form through which a data subject cannot be identified even if the information is combined with other information, after reasonably considering factors such as time, cost and technology.

3.17 **Aggregation** - refers to the process of creating higher level data by combining data across data subjects so that it does not reveal any personally identifying information about a data subject. For example, individual data that is clubbed by geographic area, by year, by service agency, or by any other attribute of a data subject that is amenable to aggregation. Aggregation is a way of anonymizing data.

3.18 **De-identification or pseudo anonymization** - means the process by which identifiers from personal data may be removed, or masked, or replaced with such other fictitious name or code that is unique to an individual but does not, on its own, directly identify the data subject.

4 OBJECTIVES

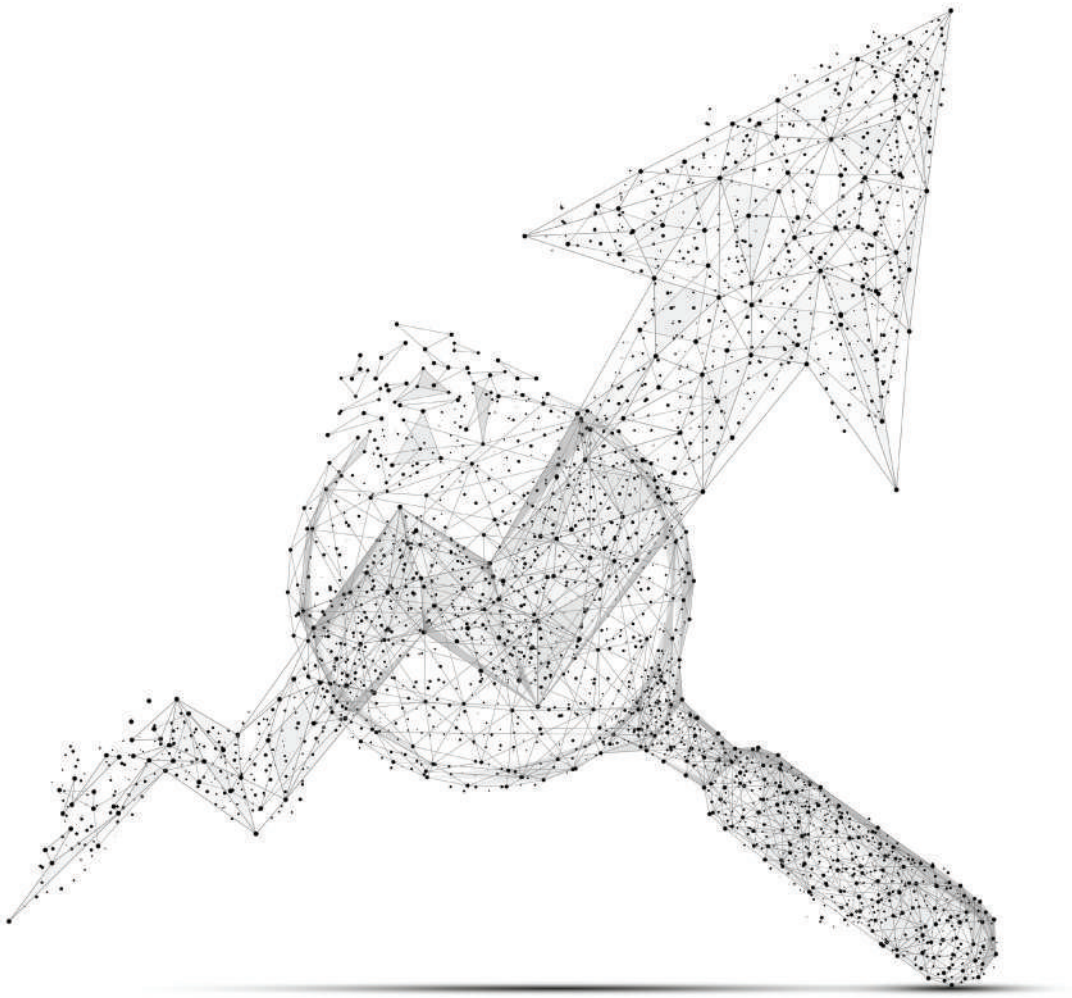
The State level data policies that have come in India have been formulated primarily aiming to improve data quality, governance and prompting all departments to publish data in a machine-readable form. While the intentions are laudable, it is not easy to quantify these objectives and benchmark where the State stands. Government of Tamil Nadu aims for outcome-based objectives to improve the public service delivery and enable evidence-based policy formulation, while maximizing data sharing to improve transparency.

Specifically, the objectives include to:

- Develop data capabilities to address all phases of governance, namely - policy making, administration, scheme review, policy redesign, aimed at developing a data-driven decision-making culture.
- Promote the data analytics capability of all departments.
- Minimize exclusion/inclusion errors in government schemes.
- Data driven assessment of efficacy of schemes.
- Making data publicly available for policy research and for improving transparency.

This will require -

- Making protocols for sharing non-aggregate, but anonymized data for research use in specific situations with safeguards.
- Prescribing a comprehensive Meta - Data Catalogue for all types of data stored in Government systems with clear ownership and definitions.
- Prescribing a data architecture that will permit and encourage cross-department data exchange.
- Effectively addressing privacy and confidentiality concerns while sharing data with non-government parties.
- Setting standardised processes to make sharing streamlined and sustainable.
- Classifying different type of data and evolving appropriate access protocols for sharing them.



5 SCOPE AND APPLICABILITY

The TNDP shall be applicable to all the Public Authority as defined under section 2(h) of the Right to Information Act, 2005 (Central Act 22 of 2005), within the State of Tamil Nadu.

The TNDP will apply to all data and information created, generated, collected, and archived using public funds of Government of Tamil Nadu directly or through authorized agencies by various Departments / Organizations / Agencies and Autonomous bodies.

This Policy will also apply to data that is recurring in nature and generated owing to automation (result output of their service delivery to citizens / business) of State user department process through various IT systems and to legacy data that is still available in non-machine-readable form.

CHALLENGES

6 CHALLENGES

Government's systems have grown to large and disconnected systems across different departments. A few examples of operational challenges are given below that we are aiming to solve through TNDP:

- Government systems capture citizens' name in different formats. Examples include name fields – first name and last name. Some capture this in a single field, some capture first name, middle name and surname, or initials, some allow spaces in each of these fields. So, the same citizen will be represented in multiple ways in multiple databases. This can lead to accidental or purposeful duplication. Most databases lack a family identifier, and this is a problem for family-based entitlements.
- There are different kinds of addresses – permanent, residence, communication addresses. For different purposes, citizens give different addresses. For e.g. in school and colleges, most give permanent address. For banks, people give permanent communication address where someone known to them is always there. For electricity connection/health, resident gives the current address. Hence, which database is the single source of correct address for a citizen is not usually defined.

- Some databases support multi-language fields or names, addresses and lot of other attributes. The fields that support multiple languages is not consistent across multiple databases.
- Numeric and text fields are not consistent across multiple databases. Different media (audio, video, social media) constrain the ability to cross reference.
- Many citizens face exclusion errors in some schemes and these could potentially be redressed by cross referencing data in other schemes to ensure that citizens are not denied benefits, subsidies and services they are entitled under the law.

The Personal Data Protection Bill, 2019 (PDP) that aims to regulate data production and sharing is yet to become a law but certain aspects of the Bill like personally identifiable information and sensitive personal information will be a factor to consider when we start using data produced in the Government to solve the above challenges.

7 BENEFITS

7.1 **Maximise use of data** - Government data has been collected to provide good governance and administration. However, the Government has not achieved its full potential due to various constraints. By providing easy access to government owned data and related metadata, citizens and other parties like public and private institutions will be able to use this valuable public resource for the benefit of the community.

7.2 **Minimise time/cost of data collection** - Government agencies and bodies collect the same data for their purposes. This entails additional effort and cost, which can be avoided if the relevant common data is shared resulting in significant time/cost savings in data collection and analysis. Standardisation of user profiles will help departmental data bases to become uniform and help in creating cross functional applications. Data can also be extracted for a beneficiary existing across different departments involving multiple departments.

7.3 **Integration** - By adhering to common standards during the collection, transfer and sharing of data, it may be possible to integrate different data sets. This would enrich the data and provide an integrated perspective to the policy maker.

7.4 **Better decision-making and inter-department co-ordination** - Data and information facilitates and improves the quality of decisions. Access to existing data is essential for many decision-making tasks that span multiple departments. Collaboration at the inter-department level for data analysis will provide rich insights gained from the data. This is critical to promote data-centric governance.

7.5 **Equitable access** - An open and transparent data sharing policy ensures access to all bona fide users. It also facilitates better data handshake between inter-department IT initiatives with common datasets but different data derivates for using data sets to render services to internal and external entities.

7.6 **Fostering Innovation** - “Data is the new oil”. Access to quality data has become a source of competitive advantage, especially in emerging technologies like AI, Machine Learning and Data Science. Quality Government data will enable predictive analysis to allow government framework to be proactive in delivery. Big corporations invest huge amounts of money in collecting and curating datasets for developing their AI-based applications. Sharing open Government data will help small companies and start-ups bring out innovative data-related products.

7.7 **Job creation** - TNDP will give a huge boost to digital firms, not only enabling start-ups to thrive alongside big firms but create new jobs in the field of data processing and analytics. Having adequate data would also give rise to multiple opportunities for research, statistical and archiving organizations, and domestic and international academic institutions.

7.8 **Minimise Exclusion error** - Citizens who may not have certain identification documents or may not be aware of benefits under other schemes could be identified and proactively provided identification/benefits.

7.9 **Protection of citizen privacy rights** - by providing a framework to protect the privacy of the residents of the State.

7.10 **Transparency** – Encourages transparency by proactive disclosure of information on OGD and on websites for the use of data by citizens/third parties.

8 RULES OF ACCESS TO NON-OPEN DATA

Access to non-open datasets of Government depends on various factors. Questions such as whether data is aggregated, who is going to use data, what data will be used for, whether data is personally identifiable, whether it has sensitive personal data are critical to determine access to the data.

The table below gives access type/cost and permissioning authority based on the sensitivity of the data set.

Data Sharing Access/Cost / Permission Matrix

Purpose of data sharing	Data Requester	Nature of Dataset				Access/ Cost	Permission Authority
		Personally Identifiable Information	Sensitive Personal Data	Anonymized	Aggregated		
Public good	Anybody	N	N	Y	Y	Open/ free	Automated
Policy making/ monitoring	Govt Depts. or Agencies/ PSUs	N	N	Y	N	Open/ free	TNeGA
Policy making/ monitoring	Govt Depts. or Agencies/ PSUs	Y	Y	Y/N	Y/N	Permissioned/ free	DIDC (see chapter 13)
Policy research / advocacy	Academia/ NGOs	N	Y/N	Y	Y/N	Permissioned/ TBD*	
Value Added Services (VAS)	Govt Agencies/ PSUs/ NGOs/ Private Sector	N	Y/N	Y	Y/N	Permissioned/ TBD*	
Any		Y	Y	Y/N	Y/N	Case by case basis	EDGC (see chapter 13)

** – to be determined under pricing policy*

Further, in all cases where data is shared with non-government research agencies and private sector, no data should be permanently retained for any purpose beyond the time permitted as per the terms of use, nor can the data be used for any commercial purpose nor can be shared with any other third parties. All data provisioning/sharing should be through secured API's.

Permissioned access to research and statistical organizations / agencies or universities of personal or sensitive personal data will require requisite due diligence, establish a bona fide research and non-commercial purpose and if the proposed output meets a public purpose/ is a public good. Such access can be granted subject to a contract executed between the data owner and the researcher that includes requirements for due diligence, periodic review and security safeguards. It also describes the mode of sharing the data; documents the non-commercial and research use on a case-to-case basis. It has a provision for penalties for any violations of the contract by data receiver. Such data, whether anonymized/personal data/sensitive personal data/disaggregated data must be available and legally shareable under applicable laws and policies.

In all cases, wherever possible, de-identified data shall be shared. However, suppose personal or sensitive personal data is necessary, such as conducting a quality survey as part of the validation of existing administrative records, checking field-level delivery of a government program, or combining with survey data on individual outcomes from welfare programmes. In that case, only minimal identifiable information shall be shared, on a case-by-case basis with required safeguards, and as is necessary for the research. Some of the purposes for which personal/sensitive personal data may be shared may include-

(i) Aid in implementing government policy;

(ii) Research and development for a public benefit;

(iii) Program design, implementation and evaluation; and

(iv) Delivery of government services.

Sharing any personal data could follow available frameworks such as the “Five Safes framework” of safe projects, safe people, safe settings, safe data, and safe outputs. The data sharing will be governed by data use agreements and non-disclosure agreements as required. TNeGA shall notify detailed process for seeking permissioned access.

Any data that can be petitioned under RTI should be proactively disclosed under Open Data policy to make the benefits available to all entities.

CROSS-DEPARTMENT ANALYSIS

9 CROSS-DEPARTMENT ANALYSIS

While open data helps in increasing transparency, key to improving governance is to use raw government digital data which has not been used so far through data analytics.

This is one of the key objectives of this Policy by promoting cross-department data analytics. This involves looking at data sets in different departments through identifiers such as demographic information, phone number, ration card, makkal number (as envisaged in the State Family Database Project). Such cross-department analysis can often give new insights for gap identification such as:

- Students who are not vaccinated against “X” disease, and are receiving scholarships.
- Families that are drawing LPG subsidies but also have 4 wheelers.
- Persons with heinous criminal cases could be restricted access to public resources.

This type of analytics is essential for the Government to develop targeted efforts to improve its benefit disbursement programs. Likewise, Open Data is useful for external research agencies to undertake policy level analysis that can result in actionable advice for the Government and provide value added services. The Government will establish an analytic layer and an anonymization layer that caters to both government entities as well as research agencies as envisaged in Figure 1.

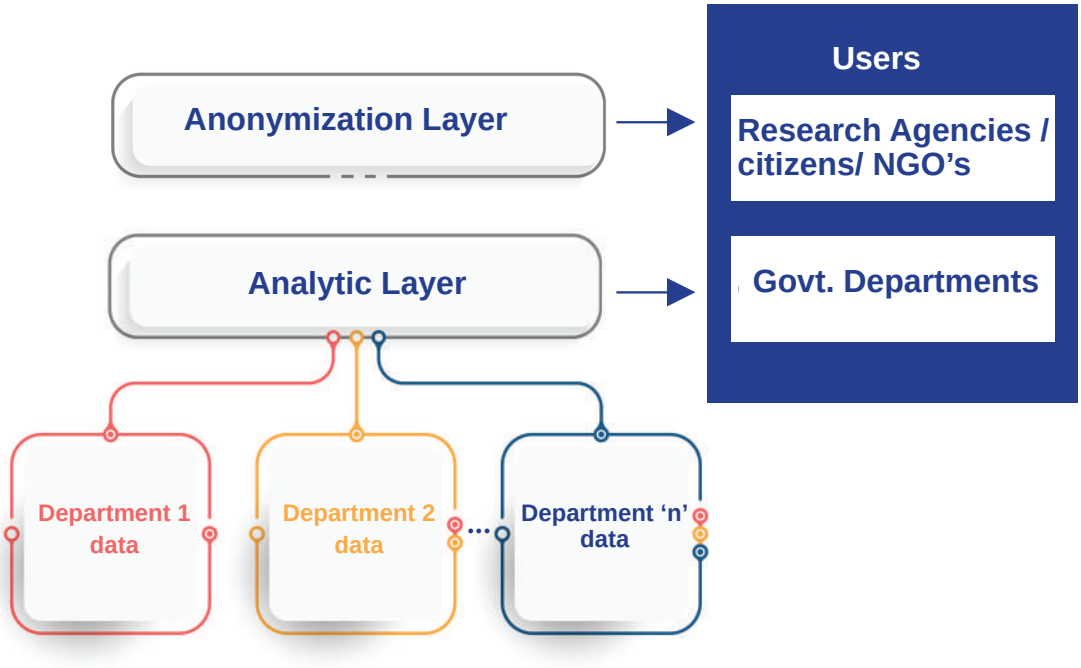


Figure 1. Data analytics framework for cross-department data analysis

Based on the data available from other States and Central Government, above data analytic framework could be expanded to include other datasets.

10 LEGAL FRAMEWORK

- TNeGA is the nodal empowered agency for Government data as per G.O.(Ms.) No.17, dated 23.09.2021 issued by Information Technology Department, Government of Tamil Nadu.
- The data standards section of **"Reference Standards"** shall be referred to for compliance with mandatory and optional data standards as and when updated/ released by Information Technology Department.
- Data collected by TNeGA will be co-owned by it on behalf of the Government of Tamil Nadu, with primary owner being the agency / department / ministry / entity which collected and provided it to TNeGA.
- The Government will legislate any additional laws and formulate rules needed to ensure adequate implementation of the TNDP.
- Access to data under this Policy will not be in violation of any Acts and Regulations of the Government of India or the Government of Tamil Nadu in force.

- Legal framework of this Policy will be aligned with various Acts and Rules covering the data as prescribed by the Government of Tamil Nadu and the Government of India from time to time.
- Information Technology Department shall be authorised to issue any operational clarifications or interpretations of the Policy and also issue amendments to the Policy as and when the need arises.

11 POLICY FRAMEWORK

The policy action and outcomes pursuant to the objectives of this Policy are given below. These apply to both open data and non-open data:

Data Standards

Actions :

- TNeGA will be the nodal empowered agency for specifying the Meta Data Catalogue and associated activities as per G.O.(Ms.) No.17, dated 23.09.2021 issued by Information Technology Department, Government of Tamil Nadu.
- A comprehensive and mandatory Meta Data Catalogue[1] for all departments will be prepared and updated from time to time which contains the definition of each of the citizen centric fields and master data, including the source of truth and responsible departments or data fiduciary.

[1] Metadata catalogue contains definitions, datatypes, length, and all other attributes of a typical data element like name, address, occupation etc. These kinds of fields are stored in multiple ways in multiple applications.

- Like State Family Database, TNeGA will create a ‘Business Registry’ comprising data of businesses available with different departments and enrich it with appropriate additional fields to enable data analytics in the G2B sphere and improve ease/cost of doing business.
- Departments sharing data will be responsible for data correctness, completeness, and authenticity.

Outcome :

- ✓ Uniform data collected in conformance with meta-data catalogue, creation of social registry and business registry, will drive good data quality which will positively impact various scheme implementation and policy research.

Data Storage

Actions :

- The State will adopt a mix of federated and centralized data storage.
- Critical master data such as the family database will be stored centrally with appropriate safeguards and protection of personally identifiable information, such as removing potentially identifiable characteristics and other statistical techniques.
- Application specific data would remain with the respective departments. Similarly, latency sensitive data will be stored near where it is needed.
- Datasets will have data retention and destruction guidelines that will be developed.

Outcome :

- ✓ Storage of data in a decentralized format will improve the processing time and mitigate the risk of loss of data.

Data Collection and Processing

Actions :

- Scheme efficiency data will be collected through a high frequency call-based monitoring cell that will be set up in TNeGA in collaboration with DEAR (Department of Economics

and Applied Research). This will include beneficiary feedback survey, data quality validation survey and outcome tracking based on departmental requests. The preferred way of collecting data would be through structured electronic survey forms with appropriate restrictions on different fields (e.g. Name / Date / numeric fields in a specific format as per the metadata catalogue).

- Departments shall implement an electronic (mobile/web based) feedback mechanism on collecting citizen ratings of the delivery of their schemes, whether the scheme is entirely physical or digital or a combination thereof, with the help of TNeGA. This data shall form the basis for evaluating citizen satisfaction of the scheme and the basis for improvements and rating schemes based on citizen feedback ratings.
- Data policy will enable a mechanism so that basic demographic information is sought only once for all schemes. This will ensure single source of data for attributes such as Name, Date of Birth, Address, e-KYC authenticated bank account, photograph, and lead to significant cost/time savings. This can also enable single change in one place that can propagate in all federated databases and help avoid rework for the citizens.
- TNeGA will publish separate guidelines for departments to follow in collecting the data during beneficiary registration and benefit delivery. Departments will be required to collect and provide data as prescribed by TNeGA.
- Once data is collected by TNeGA, before it is pushed to storage, appropriate checks will be done to ensure that data in the correct format is stored.
- Departments should integrate other data collected with service delivery, wherever possible and for purposes in this Policy to help in policy/scheme design, scheme implementation, evaluation, innovations and mid-course correction.

Outcomes :

- ✓ Reduce time/cost in enumerating beneficiaries and reduce their inconvenience.
- ✓ Reduce inclusion and exclusion errors.
- ✓ Development of indices such as the data policy effectiveness index based on inclusion and exclusion errors.

Data Sharing and Publishing

Actions :

- OGD (Open Government Data) datasets[2] will have a defined frequency of updating defined in discussion with the respective department.
- Data shall be published in machine readable formats such as csv, xml, json to minimize the use of PDFs.
- TNeGA will come up with Data API's which can be released for external developers to tap the OGD datasets.
- An analytic layer and anonymization layer will facilitate cross-department data analysis.
- All cross-department analytic data which is anonymized will be published through OGD platforms.
- All data of all departments, including beneficiary/customer/citizen/business and transaction/or any other data relating to them shall be shared with TNeGA by all departments as per G.O (Ms.) No.17, dated 23.09.2021 issued by Information Technology Department, Government of Tamil Nadu.
- Sharing of data with third parties which is not published under the OGD initiative shall have to be authorized and under suitable restrictions by Empowered Data Governance Committee (EDGC) and at a price, if any, as determined by applicable data pricing policy of TNeGA (to be framed).
- Private sector will be encouraged to come up with value added solutions using shareable data.

[2] https://data.gov.in/ . The idea of Open Government Data is that all personally non-identifiable and aggregate data should be made available in machine readable modes through multiple modes of access (csv, xml, json, api) so that third party applications can be developed using them.

Outcomes :

- ✓ Research agencies and non-profits can analyze to help in policy development.
- ✓ Dashboards based on the Government data will enhance transparency for citizens and ease of decision making for internal clients.
- ✓ Value-added services based on Government data will improve Governance, economic activity and jobs.

Data Quality

Actions :

- All data collection modes of Government will be digitized with appropriate data quality control mechanisms. Old non-digital data sets such as CRS that are critical to decision making and beneficiary selection and improving accuracy of benefit schemes shall be digitized.
- TNeGA shall prescribe automated data-quality checks for critical datasets.

Outcome :

- ✓ Continuous data quality checks ensure that benefits disbursements, scheme administration, evidence-based policy development remains efficient.

Evidence informed decision making

Actions :

For use in policy design, scheme rollout, and redesign -

- Engage with Planning (DAU), Finance, and line departments to share results from the analytics to generate insights.
- Engage with Government’s research partners to enable secure data access and integration.

Outcome :

- ✓ Social welfare schemes that are relevant and create maximum impact for people of the State.

Data Security and Privacy

Actions :

- All State departments and government organizations shall abide by all the latest data security and protection policy[3] guidelines and frameworks of the Government of Tamil Nadu, TNeGA instructions and any other rules / guidelines by the Government of India.
- Any cloud deployments of any Government centric application will be only done in MeitY empaneled cloud providers.
- Personal data of citizens shall be collected only to the extent necessary for the relevant scheme or program. Sensitization will be carried out for the Government staff handling such data.
- Personal data shall not be erased once some subsidy or benefit is availed from Government schemes.
- An audit log of central databases containing personal information would be developed to log the identity of the person/entity that accessed the personal data of an individual and the purpose for access.

Outcome :

- ✓ Ensuring that the privacy and security of citizen's data will develop people's trust in the system and administration.

[3] https://www.meity.gov.in/data-protection-framework

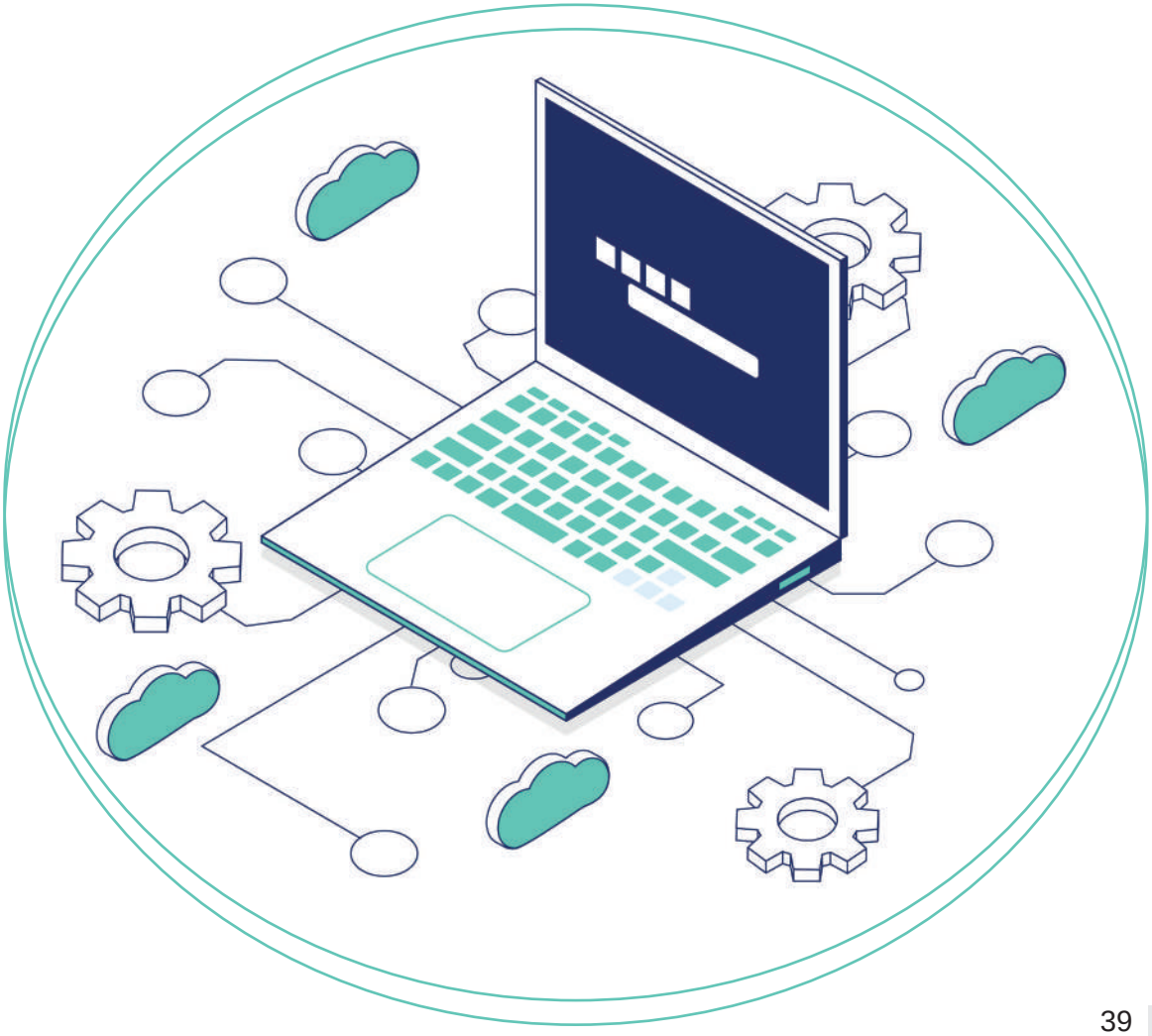
Data Governance

Actions :

- Every department and State Government Enterprise will appoint a Chief Data Officer responsible for data centric initiatives, coordination with TNeGA, compliance to data policy, OGD data, data privacy and sharing of data.
- The TNDP will be periodically amended to incorporate any new standards adopted by the Government of India and the Government of Tamil Nadu.

Outcome :

✓ Strong involvement of all departments as well as periodic updating of data policies would ensure that regular changes required in implementation programs would be done effectively.



12 PRICING OF DATA

- Data is valuable and its value is increased manifold if it is error-free, reliable and can be put to appropriate use. The TNDP aims to encourage the dissemination of and use of data freely by sharing non-sensitive data as and when possible.
- However, data collection, storage, and maintenance require significant investment on the part of the Government. Hence, for sustainability, monetization of data without compromising on equity, privacy and principles of open data is required.
- The price of non-open data to be shared, if any, would be as per the policy of the Government of Tamil Nadu and TNeGA shall be responsible for issuing instructions on data pricing.

IMPLEMENTATION FRAMEWORK

13 IMPLEMENTATION FRAMEWORK

The TNDP is mandatorily applicable to all departments/agencies of the Government of Tamil Nadu. Implementing all the aspects of data policy requires concerted efforts by all departments. To benefit from the TNDP, the departments of the Government of Tamil Nadu should undertake the following activities to bring the impact of data onto policy as shown in Figure 2:

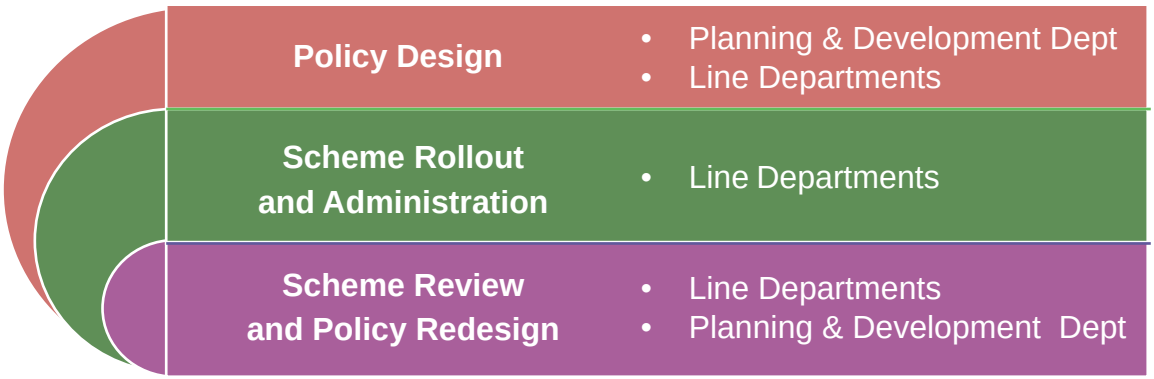


Figure 2. Responsibilities chart for the various departments

Organizational Structure

The State-level **Empowered Data Governance Committee (EDGC)** will provide strategic guidance for the State data policy framework.

This Committee would be chaired by the Chief Secretary with Principal Secretary to Government, Information Technology Department as its Member-Secretary and the Secretaries to Government of Planning, Development, Special Programme Implementation, Finance, Human Resource Management, Public and Higher Education Departments shall be permanent members of the Committee. Secretaries of the department whose data sets are being debated/considered shall be Special Invitees.

This Committee would -

- Approve exceptions not envisioned in this Policy (covering data creation, storage, sharing (including with third parties), usage, pricing, publishing, and adherence to data standards).
- Accord approvals on publishing and usage of data for high level initiatives and review data policy compliance periodically.
- Recommend to the Government, any corrective and preventive measures to be taken up to uphold the objectives of TNDP, based on reports and recommendations shared by the Data Inter-Departmental Committee.
- Approve changes/updates to the TNDP brought to the Empowered Data Governance Committee (EDGC) by the Information Technology Department.
- Appellate action on any grievances that reach them from State user departments, citizen availing social welfare schemes and any State user department services against the decision of the Data-Inter-Departmental Committee (DIDC).
- Review Data Privacy measures from time to time and take action to create/ improve the legal, technical and administrative data privacy framework.

The Data Inter-Departmental Committee (DIDC) will take operational level decisions emanating from the approved State data policy. It will be chaired by the CEO, TNeGA who shall be the State's Chief Data Officer (CDO). The Committee will comprise of

permanent members from major data generating departments as notified by the Information Technology Department. However, all departments will have to nominate one Data Officer not more than one level below the Head of Department who will be empowered by the HOD to represent him and take decisions on his behalf and as needed to participate in the committee proceedings as and when invited in relation to their data.

The Data Inter-Departmental Committee (DIDC) would be responsible for the following under the approved Data Policy framework:

- Report its activities to the Empowered Data Governance Committee (EDGC) at least once in a quarter.
- Identify the schemes for which inclusion and exclusion errors can be minimized in consultation with departments.
- Identify the problem definitions to be addressed using cross-department analytics for fixing leakages.
- Prepare a Negative List of datasets from each department within a fixed time frame.
- Prepare a schedule of datasets to be released in the next year.
- Extend Technical Support in preparation of datasets and in conversion of formats and digitization to the departments.
- Monitor and manage the Open data initiative in various departments.
- Ensure quality and correctness of the datasets through appropriate mandates to the departments.
- Permit access to data for policy making/monitoring to Government Departments/ PSUs/Government Agencies.
- Promote proactive public dissemination of datasets.
- Institutionalize the creation of datasets in various departments.
- Work out the processes and technologies (recommended or mandated) for data storage, data handling, data access, processing and use.

- Publish guidelines from time to time on how data may be shared and used for different purposes as per policy including data privacy measures.
- Make Policy change recommendations to the Empowered Data Governance Committee (EDGC).

13.1 Chief Data Officer (CDO) :

The CEO, TNeGA shall be the Chief Data Officer of the State. Responsibilities of the Chief Data Officer include:

- Lead all the data initiatives of the Government of Tamil Nadu.
- Head DIDC, which helps in compilation, collation, conversion, and publishing catalogs/ resources on the platform.
- Help Departments in minimizing inclusion and exclusion errors in schemes.
- Run cross-department analytics initiatives to help minimize leakages.
- Ensure data privacy of citizens while sharing data through all possible appropriate safeguards, but not limited to non-disclosure agreements, technical mechanisms, advisories, training of staff and awareness.
- Own the critical data quality efforts such as Meta-Data Catalogue and call center for data quality improvement.
- Define an appropriate process for the identifying and releasing datasets on the Open Data Portal.
- Prepare the Negative List for Departments in consultation with them.
- Ensure that the datasets being published on OGD follow TNDP through a predefined workflow process.
- Release as many datasets as possible on OGD proactively in consultation with departments.
- To streamline the contribution of datasets from offices/organizations for the open data under the State departments, seek several sub-Data Officers for each department.
- Work with DAU / Line departments to help them convert data into actionable insights based on their requirements.

- Implement a grievance redressal mechanism under issues arising out of the Policy for citizens and departments and take appropriate action on the feedback/suggestion received from citizens.

13.2 Data Inter- Departmental Committee (DIDC) :

To implement TNDP, each Department would nominate a Data Officer not lower than one level below the Head of Department to the Data Inter-Departmental Committee (DIDC) to be responsible for both open and non-open data.

Its work shall include the following streams:

Open Data Stream

- Identify datasets that can be shared on OGD under the TNDP.
- Identify the legitimate and legal use of the datasets published in the State Open Data Platform. It will demarcate the uses prohibited for the shared public data. For example, the shared open data may not be combined with other data, obtained privately or otherwise to extract Personal Identifiable Information (PII).
- Prepare Negative List of Datasets that cannot be shared due to legal restrictions or policy decisions.
- Publish Catalogs and Resources (Datasets/APIs/Apps) on the Tamil Nadu Open Data Platform[4] .
- Create an action plan for the regular release of data sets on the Tamil Nadu Open Data Platform.
- Monitor and coordinate the open data Program of the Departments.

[4]https://tn.data.gov.in/ (Open Data portal of Government of Tamil Nadu, where all entities under Government of Tamil Nadu can upload all the open data sets)

Data Quality Stream

- Develop and institutionalize the Meta-Data Catalogue as an ongoing initiative.
- All departments shall mandatorily consult TNeGA with respect to the IT systems being deployed, especially for data fields being collected to ensure standardization of data, avoiding duplication of basic beneficiary data and interoperability.
- Work on data quality and refinement by various survey methods.
- In consultation with line departments, identify schemes / mechanisms through which inclusion, exclusion errors can be minimized.

Analytics and Evidence based Decision Making Stream

- Develop the problem definitions that could be addressed by inter or cross-department analytics.
- Define the required data and program to manage those initiatives by institutionalizing data sourcing, quality, reporting and analytics.
- Develop the analytic layer for reporting and visualizations.

Compliance and Governance Stream

- Conduct a periodic audit of compliance with TNDP including the adherence to data standards and other guidelines issued from time to time. Report the compliance and any recommendations to the Empowered Committee.
- Act on grievances from State user departments, citizen availing social welfare schemes and any State user department services.

The **Data Inter-Departmental Committee (DIDC)** shall be assisted by professionals conversant in data analysis, portal development, database development, visualization, analytics, and programming.

Managing partnerships and Data Sharing Stream

- Be the nodal team to manage the partnerships with NGOs and research organizations.
- Evaluate the proposals for data access for both ‘public good’ and ‘value added services and allow access as prescribed in the Policy.

- Obtain the consent and necessary data protection undertakings / agreements from all data users for the intended use per their proposal.

13.3 Data Officers and Sub-Data Officers :

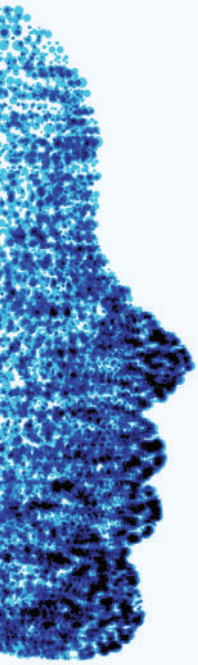
The Department Data Officer shall be responsible for its department data collection, decisions relating open data, storage, compliance to the TNDP and carry out the decisions of the **Data Inter - Departmental Committee (DIDC)**.

The Sub-Data Officers would be responsible for identifying and contributing datasets to the Tamil Nadu Open Data Platform along with their metadata. Each Sub-Data Officer shall contribute data as per the defined metadata format to facilitate easy sharing and collaboration. Prior to the contribution, the datasets shall have to be approved by the Department Data Officer. Sub-Data Officer could be an officer of the Department responsible for their unit/division. All Sub-Data Officers shall report to the Data Officer under the TNDP for all purposes.

The responsibilities of the Sub-Data Officer would include:

- Defining which data sets can be shared on Tamil Nadu Open Data platform / Sate level Central repository.
- Sharing data framework that derives appending of data after one time creation, share archival policy and data audit plans, time driven to refine data and reject redundant data.
- Responsible for ensuring quality and correctness of datasets of his/her unit/division.
- Preparing and contributing the catalogs and resources along with the metadata on the OGD Platform.

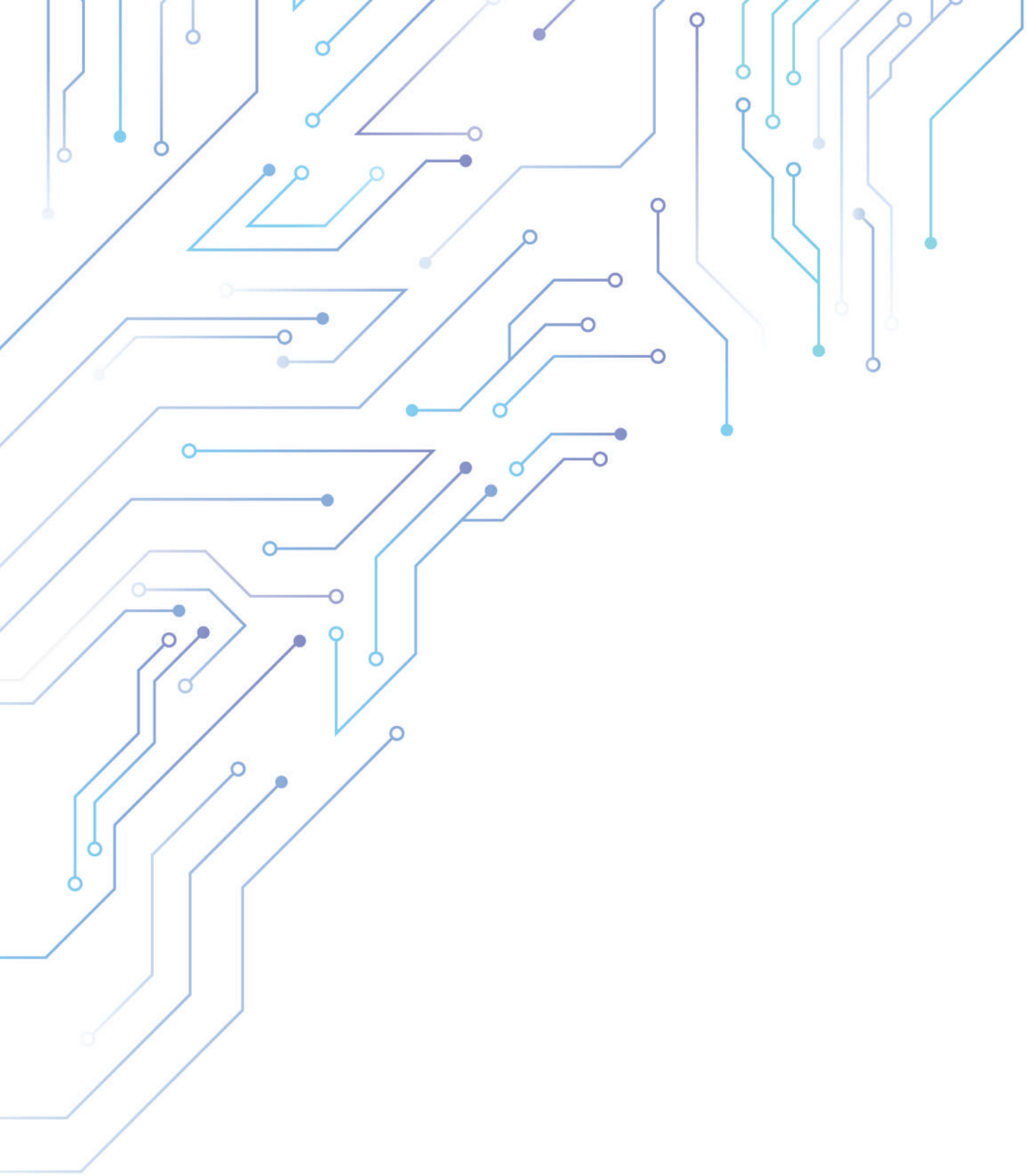
14 CONCLUSION



The TNDP aims to provide a policy framework to guide the Tamil Nadu Government Departments to use the data that multiple departments own in the best ways possible so that maximum benefit comes to the State and people. It will also ensure that the benefits and schemes of the Government are disbursed most efficiently by minimizing inclusion, exclusion errors and driving a culture of data-driven policy making.

TNDP also aims to nurture and promote of a technology-based culture of data management, data sharing and access. It makes available proactively, information on available data, which could be shared with civil society for developmental purposes.

The TNDP is a dynamic document that would need to adapt to changes in society and technological advancements in data creation, management, sharing and associated challenges. The Information Technology Department, Government of Tamil Nadu will continue evolving the policy further, keeping in tune with the technological advancements and the State's requirements.



For Information and enquiries:

Principal Secretary to Government of Tamil Nadu.
Information Technology Department
Secretariat, Chennai - 600 009, Tamil Nadu, India
Tel : 91-44-25670783 | Email : secyit.tn@nic.in
Website : https://it.tn.gov.in