Punjab State Data Policy 2020

Department of Governance Reforms & Public Grievances

Government of Punjab

Message by Capt. Amarinder Singh,
Hon'ble Chief Minister, Government of Punjab

Punjab is known for its dynamism and for spearheading changes in governance and the socio-economic ecosystem. Over the past few years, progress in technology has ushered in a new era of data-driven governance. Open data, transparency, data privacy, and e-governance have emerged as the pillars of innovative decision making across the world. The State of Punjab has been cognizant of this and has made significant progress with regard to using data and technology for effective governance. The Punjab State Data Policy is a pathbreaking move to accelerate the efficient and optimal use of data for governance.

Punjab is one of the first states in the country to formulate a comprehensive Data Policy. Through this policy, we hope to leverage data as a powerful tool for socio-economic development of the state, thereby nurturing a culture of data-driven and evidence-based decision making.

The Punjab State Data Policy shall serve as a guiding instrument to achieve the goal of promoting inclusive development and enhancing the quality of life and services for the citizens of Punjab. I am confident that this policy will act as the blueprint for the growth of a digitally empowered Punjab and set a benchmark for other states to emulate.

Foreword by Ms. Vini Mahajan, IAS,
Chief Secretary to the Government of Punjab

The rapid development of technology has heralded a paradigm shift in the way governments function worldwide. This need and pace of change is accelerated by tremendous increase in the volume of data and ubiquity of mobile devices. Growing societal reliance on digital technologies is not only reshaping citizens' expectations but also redefining governance mechanisms.

The Government of Punjab has taken several steps towards e-governance and is increasingly using data to improve service delivery and empower citizens. Taking these endeavors further and becoming one of the first States in the country to take such an initiative, the Government of Punjab has formulated a comprehensive state-level Data Policy. This overarching policy, situated within the context of broader global and national focus on data, is a significant step towards building a digital and data-driven governance ecosystem.

The Punjab State Data Policy 2020 (PSDP) shall lay down the fundamentals of data governance. It shall set Punjab on a path to embrace technology and become agile in response to ecosystem developments while ensuring citizen privacy and data security. While the detailed guidelines for applications and systems shall be issued later, this Policy shall serve as a guiding principle for all departments to collect, process, manage and use data for efficient, effective, transparent and accessible service delivery to citizens.

The Government of Punjab adopts the Punjab State Data Policy with the hope that it will lead to improved governance and citizen satisfaction.

Contents

I. Preamble

Reliable data is essential for efficient governance and transparent, effective, and accessible service delivery. Punjab has taken strides in digitization of data and is now moving to institutionalize a framework for data collection, sharing, storage, and usage. Recognizing the potential utility and importance of the data generated by the government departments, organizations, and institutions of the state for good governance, the Punjab State Data Policy (PSDP) seeks to define the rules of engagement with regards to all aspects of data management while remaining committed to protecting the privacy of citizens and making optimal use of data for evidence-based decision-making.

The Department of Science and Technology, Government of India has issued the "National Data Sharing and Accessibility Policy (NDSAP)" vide Gazette of India Notification No.II, Part 1, Sec 1 on 17th March 2012", the "National Map Policy (2005)" and the "Remote Sensing Data Policy (RSDP2001 and 2011)" which spell out data sharing principles of information generated using public funds.

The Government of Punjab adopts the "Punjab State Data Policy (PSDP)" to facilitate easy access and sharing of such Government-owned data, in an open format, for supporting sustainable and inclusive governance and effective planning, implementation and monitoring of developmental programmes, managing and mitigating disasters and scientific research aiding informed decisions, for the public good. Moreover, the PSDP shall be cognizant of principles of Non-Redundancy, Openness, Flexibility Standards, Interoperability, Quality, Efficiency, Accountability, Intellectual Property Right (IPR), Right to Privacy and Right to Information.

Taking into consideration the amount of data stored and used by the government to carry out its daily operations, especially welfare-oriented benefits to citizens, the data policy of the state shall lay down the principles of storing and sharing all kinds of data held by the state government, including personal, transactional, institutional (such as businesses, hospitals, schools) and process data (including but not limited to finances incurred, resources and institutional support put into service delivery and other details such as time, operations, etc.). This policy shall also provide to enable interoperability of data use cases for efficient use of existing data such as to minimize duplication of data entry, leverage data maintained between various departments, and ensure leakages and loopholes in service delivery are plugged.

II. Definitions

For the purposes of this policy, unless the context otherwise requires:-

1. **Cloud Technology**
   Internet based computing in which large groups of configurable remote servers are networked to allow sharing of centralized data storage, data processing tasks and on-demand access to computing resources.

2. **Data**
   Representation of information, numerical compilations, and observations, documents, facts, maps, images, charts, tables and figures, concepts in digital and/or analogue forms. This will also include other forms of data that shall be generated and/or collected by the government in the future, for example facial recognition data, etc.

3. **Data Generation**
   Initial generation/collection of the data or subsequent addition of data to the same specification. It may be categorized into primary data acquisition, satellite, and aerial by designated agencies and secondary data collection under the same specification (refers to modification or addition to the primary/initial data sets) through ground survey/data acquisition. This also covers non-human data collection via drones, bots, and other emerging internet of things (IoT) devices.

4. **Dataset**
   A named collection of logically related features including processed data or information.

5. **De-identification**
   The process by which a data processor may remove, or mask identifiers from personal data, or replace them with such other fictitious name or code that is unique to an individual but does not, on its own, directly identify the individual.

6. **Geospatial Data**
   All data which is geographically referenced.

7. **Information**
   Processed data.

8. **Machine Readable Format**
   A machine readable format (e.g.:- csv, json, xml, etc.) is a structured data format that computers and software applications can easily and automatically recognize, read and process.

9. **Master Data**
   The data which has clear ownership and doesn't change frequently and describes the core entities of the state/organization. Master data includes a consistent and uniform set of identifiers and attributes. For example, districts in Punjab, departments in the government, etc.

10. **Metadata**
    The information that describes the data sources and the time, place, and conditions under which the data was created. Metadata informs the user of who, when, what, why, and how data were generated. Metadata allows the data to be traced to a known origin and a known quality.

11. **Negative List**

Non-sharable data as declared by the departments/ organizations; personal data and the datasets which are confidential in nature and/or are in the interest of the country's security in not opening to the public.

12. **Open Data**

A dataset is said to be open if anyone is free to use, reuse, and redistribute it. It should be machine readable and easily accessible.

13. **Personal Data**

Data about or relating to a person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute or any other feature of the identity of such 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.

14. **Raw Data**

Raw data is data from the primary source. The processing of raw data results in 'information', such as a statistical table, graph, search result, etc.

15. **Sensitive Personal Data**

Sensitive Data refers to personal data or information of a person which consists of information relating to:

i.   Password
ii.   Payment instrument details including bank account or credit card or debit card or other
iii. Physical, physiological and mental health condition/s
iv.   Sex life and sexual orientation
v.   Medical records and history
vi.   Biometric information
vii. Genetic data
viii. Political opinions
ix.   Religious and philosophical beliefs
x.   Any of the above information as provided to third party for providing service
xi.   Any of the above information received from third party for processing, stored or processed under lawful contract or otherwise

16. **Shareable Data**

Those data not covered under the scope of negative list and non-shareable data.

17. **State Data Center**

The State Data Center for the Government of Punjab refers to the facility owned and managed by the Department of Governance Reforms and Public Grievances to house applications and data in SAS Nagar, Punjab, and any other such data centres and cloud services approved by the Government of Punjab and/or Government of India.

18. **Transactional Data**

Transactional data is the data that emanates from transaction of information, finances, or service.

III. Vision

The Vision of the PSDP is a State that has been transformed into a digitally empowered society where the government leverages data and technology to optimize service delivery and citizens' access and participation is promoted via user-friendly and secure e-governance systems, and data is used for evidence-based policymaking and programmatic decision making.

1. Need for Data Policy

In recent years, the Government of Punjab (GoP) has taken strides towards digitizing and storing data related to the delivery of government services across policy areas. Increasingly, this data is being used by planners and decision-makers to track progress and optimize service delivery and empower citizens to access services with greater efficiency. These efforts, situated within a broader national movement towards greater data-use, are important steps towards digital, data-driven governance.

There is a need to ensure equitable, productive, and sustainable development planning as well as good and inclusive governance. The sharing of relevant data is imperative to facilitate effective and transparent governance. Lack of appropriate standards and interoperability of data and information systems leads to ineffective and suboptimal management and practices. This calls for the adoption of data policy in the state which conceptualizes provisions for appropriate data management methods encompassing the entire continuum of data generation, handling, warehousing and accessibility by defining data standards and processes as guiding principles for all Departments and institutions falling under their purview.

2. Scope

The "Punjab State Data Policy" shall apply to all data and information created, generated, collected, and processed using public funds provided by the Government of Punjab, Central Government funds, and also International donor organizations, directly or through authorized agencies by various Departments/ Organizations/Agencies and Autonomous bodies of the Punjab State. All such data shall be stored and maintained in the State Data Centre (SDC), SAS Nagar with adequate data management/security and with systems and procedures in place for access and use by all stakeholders in the State as defined under this policy.

3. Benefits of the Policy

a. **Maximizing use**
   Access to government-owned data will enable the more extensive use of the information to deliver services and facilities in an efficient and transparent manner.

b. **Interoperability to avoid duplication**
   By facilitating the sharing of required data across departments, the need for separate bodies to collect the same data will be minimized. This will result in significant time and cost-saving in data collection.

c. **Maximized integration**
   Adoption of common standards and best practices across various datasets and MIS would facilitate the integration of individual datasets.

d. **Ownership information**
    Dedicated owners for each dataset and clear responsibilities for implementation and maintenance of datasets shall provide for data digitization, accuracy, maintenance and updation.

e. **Better decision making**
    Availability of data will enable data-driven decision making along with avenues and platforms for monitoring, reporting and planning.

f. **Better tracking for policy implementation**
    Real-time availability of data (for example those relating to weather, pollution, and traffic) will allow and open pathways for live tracking of policy implementation and impact, especially for welfare schemes.

g. **Enhance the efficiency of delivery of services**
    Data sharing and interoperability through this policy will be instrumental in reducing leakages and plugging loopholes in the delivery of welfare services.

h. **Incentivization to switch to data governance**
    The focus on insights and thus targeted impact gained from analysis and interpretation of historical data shall incentivize departments and agencies to adopt data-driven decision making.

i. **Equity of access**
    Open data practices ensure better access to all bonafide users. The PSDP shall enable the use of open access data by research and civil society organizations.

IV. Data Classification

It would be incumbent upon all the government departments to define and allow access to and sharing of all the data generated by them using government funds, subject to the following data access structure:

1. Open Access

A dataset is said to be open if anyone is free to use, reuse, and redistribute it. Open data shall be easily accessible in machine-readable formats that are optimized for machine processing. Access to data generated from public funding should be easy, timely, user-friendly, and web-based without any process of registration/ authorization. This data shall be accessible at punjab.data.gov.in as well as other websites owned by the government and/or its institutions. Following is an indicative list of the types of data which falls under Open Access:

a. **Data of aggregates**
   This includes data such as hospital beds / 1000 population, teacher-student ratio by village / 1000 population), percentage of children fully immunized by primary health centres etc.

b. **Processed/Value Added Data**
   This includes data such as GDP, per capita income, etc.

c. **Data generated through delivery of government services**
   This includes data under outcome and output indicators of delivery of government schemes and programs, e.g.- literacy rate, infant mortality rate etc. as well as process indicators on service delivery such as pendency, timely resolutions of grievances, etc.

d. **Geospatial data**
   Data consisting of primarily satellite data, maps and unit-level data for the public good, e.g.-location (GPS coordinates) and details of health facilities, sewa kendras, fair price shops, public transport routes.

No personal or sensitive data shall fall under Open Access data. Open Access data should contain low impact data i.e. wherein the loss of confidentiality, integrity or availability[1] is expected to have none or limited adverse effect on the individuals and the government department who owns the information.

2. Registered Access

Datasets which are accessible only through a prescribed process of registration/ authorization by respective departments/ organizations will be available to the recognized institutions/ organization/ public users, through defined procedures. This shall primarily consist of raw data or a database dump which has been masked for identifiers such that it can be used for research purposes by academic/research/ civil society organizations.

---

1 A loss of confidentiality is the unauthorized disclosure of information. A loss of integrity is the unauthorized modification or destruction of information. A loss of availability is the disruption of access to or use of information or an information system.

3. Non-shareable

Sensitive personal data and the datasets which are confidential in nature and are in the interest of the country's security in not opening to the public would fall in the negative list. Data which is explicitly prohibited from being shared as per Section 8 and Section 9 of the Right to Information Act 2005. The RTI Act, 2005, and Right to Privacy judgment 2017 should be taken into consideration while making this list. Few examples of non-shareable data are biometric information, medical records, etc.

Non-shareable data shall contain high impact data wherein the loss of confidentiality, integrity or availability could be expected to have a severe or catastrophic adverse effect on the government department who owns the information and individuals.

V. Data Governance Framework

The enforcement of the PSDP is critical in order to achieve the vision of a digitally empowered state. Therefore, a streamlined data governance framework has been laid down for the implementation and supervision of the PSDP. Good data governance is imperative for delivering quality services to citizens and for effective policymaking. A data governance framework will help extract value from all the data held in the state, enabling greater data access, sharing and integration at the state-level thus increasing overall efficiency and accountability. The following framework shall be adopted across all government departments and governmental organizations in Punjab:

1. Ownership for Internal Governance

The ownership of all data shall reside with the department/ agency/ institution for which the data has been primarily generated directly or indirectly. In cases where the data is being generated for multiple agencies, the department/ agency/ institution responsible for the execution and/or coordination of the project shall be the primary owner of the data being generated. However, guidelines for access to this data by all parties who require access for planning and executing the process shall be laid down by the owner department/ agency/ institution as may be needed.

For all data collected generated and/ or processed by the Government, it shall be the responsibility of the owner department to ensure that relevant stakeholders at every stage of data generation, processing, storage, and use process are sensitized to relevant data policies. The owner department shall also appoint nodal officers/ staff at each level of data reporting, digitization, processing and management to ensure adherence to policies.

2. Collection and Processing

a. **Data collection shall be conducted with the utmost integrity.** The data collected should be free of bias. The desired outcomes of the service delivery or policy implementation should not influence data entry. Further, the outcomes of the service delivery or policy implementation should not affect the incentives for the frontline workers and other government officials involved in data collection.

b. **Data collection shall be integrated with service delivery, where possible.** Frontline workers and other government officials involved in the delivery of services to citizens shall collect relevant program data as a regular part of program/ service processes.

c. **Data shall be collected electronically where possible, and paper-based data will be digitized.** Investments shall be made to transition fully from paper-based systems like registers, diaries, and forms, to electronic data collection platforms. Where paper-based data collection is necessary, regular data entry processes shall be standardized to ensure digitization of all data. It must be ensured that raw paper data is digitized without loss of information. All paper-based data collected shall be digitized by the corresponding department within twenty (20) working days of its collection. Departments may request for an extension of this timeline for specific schemes/programs/datasets by sending a written request to the State Data Steering Committee (SDSC) for approval.

d. **Basic data cleaning shall be done systematically and automated where possible.** Data shall be regularly run through automated processes and algorithms and other state-of-the-art data cleaning mechanisms to verify and validate data for quality, accuracy and/or consistency. These automated checks (such as mobile number fields should allow only 10 digits) shall be built into the data collection platforms for all departments.

e. **Input constraints and automatic data checks shall be used to standardize data entry quality.** Data entry interfaces shall be designed such that extensive use of dropdowns, select one/select multiple display options, and string validation rules is made to minimize data quality issues at the entry stage. The concerned department shall identify and implement such constraints and checks. The technical expertise of DGR&PG may be leveraged for the implementation of such checks.

f. **Data collection and storage shall be compliant with data privacy and security laws and best practices.** Data collection shall be ethical and data security measures such as encryptions, firewalls, and access controls shall be established to prevent unauthorized access to confidential data, data manipulation and data loss.

3. Quality and Relevance

a. **Data shall be collected according to decision-making and reporting needs.** Data is responsive to decision-makers' monitoring, reporting, and planning needs. The collection of data that does not serve a clear purpose shall be minimized.

b. **Data shall be regularly checked for quality.** Purposive or random samples of data shall regularly be run through a standard list of data quality checks, and in-person supervision helps validate data quality. The department shall enforce these checks for all datasets in its purview every six (6) months.

c. **Datasets shall be updated.** Updates shall occur on a real-time basis by digitization of data collection. Where data is collected on paper, digitization shall be completed within twenty (20) working days of collection of such data. Departments may request for an extension of this timeline for specific schemes/programs/datasets by sending a written request to the Administrative Secretary, Department of Governance Reforms and Public Grievances for approval. The State Data Steering Committee (SDSC) shall be informed of the request and the final decision by the requesting Department.

d. **Appropriate control mechanisms to keep a check on data manipulation** shall be implemented at each level, such as the provision of password protected edit rights to specific users only, maintaining a log of access, entries and changes made to records to trace the user, date and time of the entry/change, etc.

e. **Collaborations for independent data audits for datasets** where data is generated from the information entered into the system shall be done by the departments who own the information being entered.

4. Interoperability

The government department(s) and government organization(s) of the State of Punjab have a rich and diverse repository of various kinds of data, which may overlap from duplication of data entry and/or generation. Additionally, data collected, generated and/or processed by one department may be required to improve service delivery and governance in another department. There is, therefore, a need to enable streamlined interoperability of data systems between the departments in order to avoid duplication, redundancies and leakages in service delivery. However, while the data structures, infrastructure and systems shall be designed and built to enable streamlined interoperability, the access to any dataset owned by one department/ agency by another department/agency shall be exclusively on a need basis. Departments/ agencies may submit their proposition for such requests to the SDSC for consideration and approval such that the Chief Data Officer may coordinate for its implementation.

The following may be taken as guiding principles to build mechanisms for data interoperability within and across departments and organizations in the state government

a. **Metadata repositories shall be built, published, and updated regularly.** All government departments and organizations shall invest in building and maintaining metadata repositories and publish these to improve awareness of what data exists and whether and how datasets across geographies/sectors/federal levels/time periods can be integrated

b. **Standardization of data objects that appear across datasets.** Regional hierarchies, population subgroups, identifying data and other generic data objects that appear across datasets shall be stored using common standards and guidelines built by DGR&PG so that data from across departments can be integrated. Updates to the common schema shall be coordinated across departments.

c. **Datasets shall be stored in standard machine-readable formats.** Such formats are used by other databases and therefore reduce compatibility issues, minimize conversion effort, and enable the easy exchange of data.

5. Sharing and Access

The principles on which data sharing and accessibility need to be based include openness, flexibility, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, compliance to set standards, interoperability, quality, security, efficiency, accountability, sustainability and right to privacy. The following are general practices to ensure adherence to principles as listed above:

a. **Data shall be published in machine-readable formats with easy bulk download options.** Use of formats such as csv, xml, json etc. shall be promoted while minimizing the use of PDFs and word documents to publish data. Data across geographical and time units shall be made available for download in bulk.

b. **Published data shall be made easily searchable and consistently available.** Data platforms shall be organized in an intuitive manner so that users can easily find up-to-date metadata repositories for all published datasets. Once published, datasets shall always be available and historical data be well maintained.

c. **Data shall be made available to decision-makers via user- friendly interfaces.** Data platforms shall be designed to provide decision-makers with basic analytics (indicators and summary statistics) and visualizations (such as bar graphs and trend lines) and the flexibility to adapt default analytics to their decision-making needs. These interfaces shall be designed to update on a real-time basis, with utmost accuracy. The usage of such interfaces by decision makers shall be promoted to enhance data-driven governance and policy making in Punjab.

d. **Dashboards shall be made available for citizens to engage with real-time data with minimal access barriers.** Citizens' access to data shall be facilitated by presenting data in simple, engaging formats and by minimizing access barriers (such as captchas, confusing site navigation, or slow page load times). Citizens shall be provided with the option to download underlying data for further analysis. All Departments shall prepare and publish on their own website as well as on the State open data portal which shall be built and maintained by DGR&PG, the list of all data held by them. All the concerned departments shall maintain, update and standardize such data regularly to facilitate easy and uninterrupted access to these data by all the government departments and bona fide users as per the conditions laid down under the data policy.

6. Security and Protection

a. All state departments and government organizations must abide by all the latest data security and protection policy guidelines and frameworks of the Ministry of Electronics and Information Technology (MeitY), Government of India. These guidelines include and are not limited to -

   i. Use and implementation of state of the art monitoring tools
   ii. Security audit at the state data center by third party experts periodically (once in six months)
   iii. Periodic analysis of system audit logs to identify inappropriate and illegal access and to mitigate security threats
   iv. Implementation of data back-up and recovery processes
   v. Regular updates to all IT infrastructure

b. DGR&PG shall be the nodal department for defining, maintaining and tracking the benchmarks on data security and protection for the Government of Punjab. These benchmarks shall be derived from policy guidelines of MeitY and any additional indicators as approved by the SDSC. A department wise status report on compliance with guidelines on data protection and security shall be provided to the SDSC at regular intervals by DGR&PG.

c. The DGR&PG may notify a detailed policy for data security and protection in continuation to and in compliance of this policy.

7. Data Privacy

a. Privacy is a fundamental right, flowing from the right to life and personal liberty under Article 21 of the Constitution. Privacy of personal data and facts is an essential aspect of the right to privacy. In this regard, no personal data shall be processed by any person, except for any

specific, clear and lawful purpose. The personal data shall be collected only to the extent that is necessary for the purposes of processing of such personal data. Any processing of personal data can be done only after de-identification.

b. The usage and transfer of personal data of citizens shall abide by the Information Technology (IT) Rules 2011, under the IT Act, 2000, and any other national or state acts and/or policies which may follow and cover data privacy in its scope.

c. All department(s)/government organization(s) to note that any Aadhaar data shall be considered to be highly sensitive and must be used only as per the provisions of the Aadhaar Act, 2016, as may be amended from time to time.

d. The DGR&PG may notify a detailed policy for data privacy in continuation to and in compliance of this policy.

8. Data continuity and retention

a. Data retention systems should be implemented to manage and ensure data retention for the period necessary to satisfy the purposes of data collection.

b. Guidelines for data recovery, continuity and protection shall be prepared by DGR&PG.

9. General Guidelines

a. Every department should build a Management Information System (MIS) for the purpose of data management and maintenance. Within eighteen (18) months from the date of notification of this policy, all departments will operationalize their MIS and digitize all data held and/or managed and/or used by the department. The MIS should be reviewed by the SDSC for standardization of data across all departments in the state.

b. Detailed implementation guidelines including the technology and standards for data collection, maintenance, and processing; and metadata would be published by the DGR&PG, Government of Punjab.

c. All departments shall maintain metadata repositories. All metadata will follow standards and will contain adequate information on proper citation, access, contact information and discovery. Complete information including methods, structure, semantics and quality control/ assurance is expected for most datasets.

d. Metadata relating to geospatial datasets would contain information on scale, projection, datum, resolution and date of acquisition of sources like satellite image/ aerial photos, etc.

e. The departments shall ensure that the software used in creating and generating geospatial data and information in future complies with Open Geospatial Consortium (OGC) standards.

f. All shareable data shall be made available, by the departments, to a web-based citizen dashboard on an 'as-is where-is' basis. Masked and de-identified data dumps shall also be available on the interface for research and academic institutions.

g. All departments will identify and prepare a comprehensive list of data sets maintained by them mentioning the frequency of data updation for each dataset. Within six (6) months from the date of notification of this policy, all departments will create a structure for each dataset with metadata in a standardized format. All Departments will also categorize the datasets into non-shareable data, registered access data and open access data within five (5) months of the notification of the policy, which will be reviewed by the SDSC once every year.

h. For extension of any of the above mentioned timelines, the departments may request to the State Data Steering Committee in written for approval.

VI. Implementation Guidelines

1. Supervision and Monitoring

Entities for Supervision and Monitoring of data policy shall be set up at three levels. They shall be responsible to ensure compliance with the state data policy and other relevant policies and processes.

a. **A Stata Data Steering Committee (SDSC)** will be created chaired by the Chief Secretary to the Government of Punjab and comprising of the following members:

   i. Administrative Secretary, DGR&PG
   ii. Administrative Secretary, Department of Planning
   iii. Director Governance Reforms, DGR&PG
   iv. 3-5 officers/experts, as nominated by the Chairperson.

   The composition of the Committee may be reviewed and revised as and when needed and at least once in every 6 months.

   The Director Governance Reforms (DGR), DGR&PG shall act as the Convener for this Committee.

   The role of SDSC will be to coordinate and monitor the implementation of this policy through close collaboration with all state government departments and agencies. SDSC will also establish standards, undertake data analysis, interpretation and dissemination at the state level.

   The SDSC shall appoint a Chief Data Officer (CDO) supported by a team which shall have a relevant background in technology, public policy, economics, statistics, data analytics, law, and/or other relevant subjects to support the SDSC on implementation of this policy and related projects.

   The SDSC shall be responsible to:

   i. Advise various departments with regards to the use of data and information in the developmental planning of the State.
   ii. Notify what constitutes as master data which shall be standardized across all the departments. Approve and enforce any additions or changes to be made in the master data, such as the creation of new districts or change in nomenclature of districts, for all the departments.
   iii. Formulate future policies/ guidelines regarding classification and use of data and information in the planning process, disaster mitigation and scientific research of the State.
   iv. Promote and facilitate the sharing of data and information among various State Government Departments and other users with policy formulation and implementation focused on the development and sustainable use of geospatial databases.
   v. Review and monitor the implementation of data sharing and accessibility practices every quarter.

b.   All departments will constitute a **Department Data Cell** under the chairmanship of the Department Administrative Secretary. A **Data Officer** will be appointed by the Administrative Secretary who shall be exclusively responsible for supervising and ensuring compliance to the PSDP within the department as well as undertake practices for data digitization, management, processing, quality, and analysis for evidence-based decision making.

The Data Officer shall report to the Administrative Secretary and send the department's review report on the implementation of PSDP to the CDO every 3 months. In addition to the Data Officer, it is recommended that the department also recruit or appoint additional staff under the data cell as per requirement.

This cell shall also be responsible for sharing, publishing, and standardization of the Departments' data and for coordinating with SDSC on a regular basis.

This cell shall serve as the nodal agency for implementation of all data governance-related practices within the department for this as well as future policies relevant to governance.

c.   A **District Data Cell** shall be constituted in each district under the chairpersonship of the Deputy Commissioner to supervise, coordinate and ensure compliance to this policy at the district level as well as undertake practices for data digitization, management, processing and analysis for evidence-based decision making.

The manpower with the requisite skills for the District Data Cell in each district may be provided by the Department of Planning.

The District Data Cell shall coordinate with the Department Data Cells as well as the Chief Data Officer to ensure that all practices, as outlined under this policy, like those spearheaded by the Chief Data Officer and/or the Department Data Officers are implemented in their respective districts.

d.   In addition to these internal teams, the SDSC, under the chairpersonship of the Chief Secretary, shall constitute an **Expert Group for Data Governance** with individuals and organizations outside the government who are reputed in the fields of data management, data-driven decision making and evidence-based policy and development. This Expert Group shall review the practices, standards, guidelines, metadata and mechanisms for data digitization, quality and security to suggest improvements.

This Expert Group shall meet at least once in every 6 months with the SDSC, the Chief Data Officer and the Department Data Officers for this discussion.

2. Storage and dataset organization

a.   All departments will store their datasets at **State Data Centre** as defined under Section II-17. The DGR&PG would be responsible for complying with all the data protection and security protocols. The departments may also store their data on another cloud technology service as a backup if required.

b.   Each department shall create and maintain a single common comprehensive repository of its beneficiaries/ users which shall act as the single source of truth for the delivery of all services under the department to those beneficiaries/ users. This shall be known as the '**Citizen Register**'. The department shall ensure that this database is kept updated for efficient service delivery.

c. Each department shall create and maintain a single comprehensive repository of its employees and all staff/ workers engaged to work with the department temporarily or permanently, paid as well as unpaid. This shall be known as the '**Employees Register**'. The department shall ensure that this database is kept updated and can be appropriately integrated with other human resource management platforms in use by the Government.

d. For some entities, as listed below, a specific department shall be responsible to create and maintain a data repository related to the type of entity and all other departments shall take this repository as the single source of truth for the type of entity for all transactional and service delivery purposes:

   a. The Department of Industries and Commerce shall create and maintain a repository of all commercial as well as non-commercial enterprises registered in and/or operating in Punjab. This shall be known as the '**Business Register**'. The Department of Industries and Commerce shall ensure that this database is regularly updated and reliable for use and processing.
   b. The Department of Revenue, Rehabilitation and Disaster Management shall create and maintain a repository of all land falling within the boundaries of Punjab. This shall be known as the 'Land Register'. The Department shall ensure that this database is regularly updated and reliable for use and processing.
   c. More items under this list may be notified by the SDSC, as per requirement.

VII. Data for Decision Making

In order to encourage data driven decision making and evidence based policies, the following measures shall be taken:

1. The Chief Data Officer shall be responsible to prepare outcome based budget propositions with the Department Data Officers such that it can be used by the Department of Planning and Department of Finance for evidence based budget allocation.

2. The Chief Data Officer and the Department Data Officers shall actively engage with research, academic and civil society organizations on purposes of data analysis and interpretation that can support the department/s with insights that enable evidence-based decision making and anticipatory innovation governance.

3. Reports and dynamic dashboards shall be developed and maintained with input, output and outcome indicators to indicate the socio-economic progress made by the government under each sector/ department/ district in the state. Efforts shall be made to also make accessible data on budget allocation and expenditure for each scheme and programmatic intervention.

VIII. Training and Skill Development

Alongside a strong IT infrastructure, governance framework and guidelines for data governance, it is crucial that the government manpower is regularly upskilled on technical skills, practices and know-how of data handling, management and use such that this policy may be successfully implemented, and data may be used to its potential for evidence-based decision making. Training and skill development shall be designed to equip all government manpower with the appropriate and up-to-date knowledge and skill sets such that a culture of data-driven governance is encouraged and adopted by all in the Government of Punjab.

1. SDSC, in consultation with the Expert Group, shall lay down a plan for regular training and skill development of all government manpower, especially the manpower involved in data collection, processing, management and/or use.

2. Under the guidance of the SDSC, the DGR&PG and Department of Planning shall develop training modules which shall be made available online and can be used by all officials and staff engaged in data collection, management, processing and/or use to enhance their skills on the matter.

3. The Department of Planning shall develop, make available, maintain and regularly update content on data governance including data handling, management and use.

4. The DGR&PG shall develop, make available, maintain and regularly update content on the IT infrastructure and protocols used for data collection, processing, management and use.

5. The DGR&PG shall also build a common platform where this content can be accessed and used by all government officials and staff.

6. The Chief Data Officer under the SDSC shall be responsible for necessary coordination and execution of this exercise, in consultation with the Department Data Officers and the District Data Cells.

7. All training shall be customized to the requirement at each level of data governance and management. Special emphasis shall be given to the training of district level officers and staff involved in data collection, cleaning, handling, reporting and processing at the granular level.

8. Departments may enter into collaborations with universities, academia and other such knowledge partners for designing and implementation of skill building modules, data-driven innovations and/or other related initiatives.

IX. Continuous Improvement

1. **Periodic Review of Data Policy:**
   This policy shall be reviewed by the SDSC in coordination with the Expert Group, at least once every 2 years to update and revise the policy as may be required in accordance with the applicable laws, infrastructure and other developments.

2. **Third-party assessment of policy implementation and practices:**
   As has been mentioned under Section V-3-e, the Chief Data Officer under the guidance of the SDSC and the Department Data Officers under the guidance of their respective Administrative Secretaries shall actively build collaborations with external organizations for third party audit and assessment of datasets generated from information being entered into the MIS; guidelines being used to process and use the datasets under the MIS; and the MIS being used by each department/agency. Such assessments are recommended to be done once every year.

X. Budget Provisions

The implementation of Punjab State Data Policy is expected to entail expenditures for both data owners and data managers for data conversion, data refinement, data storage, quality upgradation, etc.

The administrative departments/public organizations may approach the Department of Finance for appropriate budget allocation in this regard.

XI. Conclusion

The State Data Policy is formulated with a mandate of using government owned data and information in the interest of citizens of the state. In line with this policy, various departments will be encouraged to share data and information with each other and other stakeholders including citizens in order to enable data-driven planning and progress tracking thereof. The Punjab State Data Policy will provide greater scope for better, data-informed decision making in the State while adhering to the highest data protection standards and commitment to the principles of Data Privacy.

The scope and objectives of this policy may be expanded and enriched in future, as the need may be.

XII. Annexure: Use Cases

A few use cases of where this policy can help improve governance and support evidence based decision making are shared below:

Use Case 1:

Integrating ICDS and Health data systems to improve coordination between Anganwadi Workers (AWW), Accredited Social Health Activists (ASHA), and Auxiliary Nurses and Midwives (ANM)

AWW, ground functionaries under the Department of Social Security & Women and Child Development, and ASHAs, and ANMs, under the Department of Health & Family Welfare, have overlapping mandates to provide maternal and child health and nutrition services, there is no formal mechanism for these three crucial frontline workers to share data and coordinate effectively. Currently, ASHAs and AWW identify beneficiaries and mobilize community members by conducting home visits, and informally but often independently share data with the ANM, who then delivers key health services like immunizations, antenatal care, basic diagnostics, treatment, and referrals.

Integrating and streamlining the data systems in the two departments used by these three workers – especially systems used to identify beneficiaries and track their outcomes – would help promote the convergence. This a major goal under the POSHAN Abhiyaan initiative, which will provide a host of benefits such as enabling better targeting and prioritization of home visits, promoting accountability for case identification and ensuring that no beneficiaries are missed, and minimizing the duplication of data collection effort.

For example, AWWs are responsible for conducting growth monitoring every month, which means they are best placed to identify children who are suffering from severe acute malnutrition (SAM). However, it is the ANM who delivers the care to SAM children. If AWWs and ANM use a shared portal where the AWW can mark SAM children identified during her round and send a notification to the ANM, the ANMs can prioritize their house visits more effectively than they are in status quo, where such data sharing does not happen, or happens in an unstructured, informal way. Similar use cases exist for other overlapping beneficiary groups (such as pregnant and lactating women).

Use Case 2:

Linking cadastral maps to ownership and cultivation data to create a geo tagged village level planning tool

The Revenue department is in the process of digitizing Punjab's cadastral maps. This could be an opportunity to create a powerful map-based planning tool that links the digital cadastral maps to the Record of Rights and mutations data, crop inspection data, and village level administrative data. This can be facilitated through interdepartmental data sharing among the Department of Revenue, Rehabilitation & Disaster, Department of Agriculture and Department of Rural Development and Panchayats. Linking these data sources, would allow a user to theoretically click on any land parcel in any village to understand the agricultural details such as the cultivation arrangement, type of soil, principal kharif and rabi crops, and irrigation system

as well as understand whether the land is currently under dispute, and potentially what the ownership structure is. Most of these data points are being captured by the local revenue officers but are currently only available in single PDFs in the digitized records of rights and the crop inspection registers.

An integrated planning tool like this could be of use to different decision makers such as someone trying to identify vulnerable villages/areas within a village, or someone trying to assess eligibility for various agricultural schemes (such as PM KISAN, where availing benefits requires some proof of land ownership). Such data, visualized at this level of detail, would also inform infrastructure planning (such as by allowing a user to assess where there is empty land to build on, or whether certain areas are under dispute, etc.)

Use Case 3:

Linking the Department of School Education's student database with the Department of Social Security & Women and Child Development and Department of Water Supply & Sanitation to assess schemes focusing on girls and people with disabilities (PWD).

The Department of School Education has a huge student database which is managed through ePunjab School MIS. This database has information on personal and financial records of each student. This data can be shared with the Department of Social Security & Women and Child Development and the Department of Health to facilitate schemes that focus on indicators and outcomes pertaining to younger and adolescent girls, and people with disabilities (PWD), such as Beti Bachao Beti Padhao, Poshan Abhiyaan, and so on. For example, the dataset on 'separate toilets for girls in schools' will be owned by the Department of School Education, and the Department of Social Security and Women and Child Development, and the Department of Water Supply and Sanitation shall have access to this dataset.

Department of Governance Reforms & Public Grievances
Government of Punjab