Draft

National Data Sharing and Accessibility Policy (NDSAP)

1. Introduction

1.1 Data are recognized at all levels as a valuable resource that should be made publicly available and maintained over time to ensure that their potential value is realized. There has been an increasing demand by the community, that data should be made more readily available to all, to enable rational debate and better decision making. Principal 10 of the United Nations Declaration on Environment and Development (Rio de janeiro, June 1992), stated "……..*each individual shall have appropriate access to information concerning the environment that is held by public authorities ……. and the opportunity to participate in the decision making process. States shall facilitate and encourage public awareness and participation by making information widely available."*

Section 4(2) of the Right to Information Act, 2005 reads "It shall be a constant endeavour of every public authority to take steps in accordance with the requirements of clause (b) of sub-section (1) to provide as much information suo motu to the public at regular intervals through various means of communication, including internet, so that the public have minimum resort to the use of this Act to obtain information"

1.2 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, Interoperability, Quality, Security, Efficiency, Accountability, Sustainability.*

1.3 There is large quantum of data generated at the cost of public funds by various organizations and institutions in the country. Most of this data, is non-sensitive in nature and can be used by public for scientific, economic and developmental purposes. The National Data Sharing and Accessibility Policy (NDSAP) is designed so as to apply to all non-classified data collected using public funds held by various Ministries / Departments /Subordinate offices. The NDSAP policy would help data users and data solicitors get access to data through established procedures and defined norms.

2. Objectives

The objectives of NDSAP are to address all issues related to data in terms of the available scope of sharing and accessing spatial and non-spatial data under broad frameworks of standards and interoperability.

a) Data Classification
b) Technology for sharing and access
c) Current legal framework (RTI Act and Privacy Act)

3. Benefits of the data sharing policy

- **Maximising use**: Ready access to government data will encourage more extensive use of a valuable public resource for the benefit of the community.

- **Avoiding duplication**: By sharing data the need for separate bodies to collect the same data will be avoided resulting in significant cost savings in data collection.

- **Maximised integration**: By adopting common standards for the collection and transfer of data more integration of individual databases will be possible.

- **Ownership**: The identification of owners for the principal data sets enable users to identify those responsible for implementing prioritized data collection programs and for developing data standards.

- **Better decision-making**: Quality information allows to make competent decisions. Avoiding large potential costs. Ready access to existing spatial data is essential for many decision making tasks such as protecting the environment, development planning, managing assets, improving living conditions, national security and controlling disasters.

- **Equity of access**: A more open data transfer policy ensures better access by all bonafide users.

4. Definitions

"Data" means a representation of information, knowledge, facts, concepts or instructions which are being prepared or have been prepared in a formalized manner, and is intended to be processed, is being processed or has been processed in a computer system or computer network, and may be in any form (including

computer printouts magnetic or optical storage media, punched cards, punched tapes) or stored internally in the memory of the computer[1].It also includes data in conventional form on paper and other media.

Sensitive personal data - Sensitive Personal data or information of a person shall include information collected, received, stored, transmitted or processed by body corporate or intermediary or any person, consisting of

- any proceedings for any offence committed or alleged to have been committed by him, the disposal of such proceedings or the sentence of any court in such proceedings.
- information related to financial information such as Bank account/credit card/debit card/other payment instrument details of the users
- physiological and mental health condition
- Medical records and history
- Biometric information
- information received by body corporate for processing, stored or processed under lawful contract or otherwise

Data set - A named collection of logically related features including processed data or information.

Data Archive – A place where machine-readable data are acquired, manipulated, documented, and distributed to others for further analysis and consumption.

Data Acquisition - Initial acquisition (collection) of data or subsequent addition of data to the same specification, including data quality assurance processes.

---

1 The Information Technology Act 2000 and amendment 2008

Data Enclave – A controlled, secure environment in which, eligible users can perform analyses using restricted data resources.

Metadata – 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. Meta data allows the data to be traced to a known origin and know quality.

Negative list – Non sharable data as identified by the ministries / departments

Raw Data – Field observations, contents of project-related data study repositories, survey results, results of laboratory studies and preliminary analysis.

Restricted Data - Datasets that cannot be distributed to the general public due to confidentiality concerns, national security considerations, or other issues.

Standards - Compliant Applications – Any application that embeds data handling functions (e.g., data collection, management, transfer, integration, publication, etc.) and operates on data in a manner that complies with data format and data syntax specifications produced and maintained by open, standards bodies.

Spatial Data - Data representing geographically referenced features that are described by geographic position and attributes. Typically it includes data about natural resources, the environment, land use, demography and socio-economic.

Unique Data – Data that cannot be readily replicated.

5. Non-shareable data (Negative List)

National security and privacy are paramount to the country and individual respectively. In view of this it is mandated that each government ministries / departments need to prepare a 'negative list'. The negative list is that which includes the data that is not sharable and the same would not be available on the public domain. Sections 8 and 9 of the Right to Information Act, 2005, The Information Technology Act, 2000 and the 'right to privacy' upheld by the Hon'ble Supreme Court of India in its various judgements, need to be consulted/taken into consideration while preparing the 'negative list'.

6. Shareable Data

6.1 The other data sets identified by the ministries / departments which have not been included in the negative list shall be verified and validated by the individual departments and then ported on the website www.data.gov.in . Each funding organization shall highlight data sharing policy as preamble in its RFPs as well as project proposal formats. The funding organizations shall clearly mention mandate for projects as far as NDSAP requirements are concerned.

6.2 Appropriate support and incentives for data clean up, documentation, dissemination and storage shall be given by funding agencies.

6.3 The metadata indicating what data is accessible from the ministries / departments shall also be ported on the data.gov.in website. The metadata should contain information related to the data sets available, their quality and the data formats.

7. Data Classification

Different types of data sets are generated by different ministries /departments. The types of data produced by a statistical system consist of derived statistics like national accounts statistics, indicators like price index, data bases from census and surveys. The geospatial data however, consists primarily of satellite data, maps, etc. In such a system, it becomes important to maintain standards in respect of metadata, data layout and data access policy.

Data sets are to be classified on various types

a) Open Access data

Open access to research data from public funding should be easy, timely, user-friendly and Internet-based.

Timing of data sharing:- Data should be made openly available as soon as possible but no later than 3 months after the data was collected.

b) Registered Access

The users are required to register their names through the web and then download the information needed using the user name and password provided to them at the time of registration.

c) Restricted Access

Access to the following categories of information, in case these are not already in public domain – are restricted:

- Exact coordinates of strategic locations; sensitive archaeological, cultural and historical locations;
- Information about persons in terms of protection of data privacy
- Protection of intellectual property rights

The data users who are accessing / using this data for research should clearly acknowledge the ministry / department in all forms of publications.

8. Technology for sharing and access

A state-of-the-art data warehouse with online analytical processing (OLAP) capabilities, which includes providing, a multi-dimensional and subject oriented view of the database needs to be created. This integrated repository will hold data of current and historical nature and this repository over a period of time will also encompass data generated by various Central Ministries, State Governments and UTs; The main features of the data warehouse need to include:

(a) User friendly interface

(b) Dynamic / pull down menus

(c) Search based Report

(d) Secured web access

(e) Bulletin board

(f) Complete Metadata

(g) Parametric and Dynamic report in exportable format

9. Current legal framework

Data access arrangements need~~s~~ to respect the legal rights and legitimate interests of all concerned stakeholders. Access to, and use of, certain data will necessarily be limited by various types of legal requirements, which may include restrictions for reasons of:

- National Security: data pertaining to intelligence, military activities, or political decision making may be classified as non shareable data.
- Privacy and confidentiality: data on human subjects and other personal data are subject to restricted access under national laws and policies to protect confidentiality and privacy.
- Trade secrets and intellectual property rights: Data access arrangements should consider the applicability of copyright or of other intellectual property laws that may be relevant to publicly funded databases.

- Protection of rare, threatened or endangered species: In certain instances there may be legitimate reasons to restrict access to data on the location of biological resources for the sake of conservation, archeological sites etc.

- Legal process: data under consideration in legal actions (sub judice) may not be accessible. Subscribing to professional codes of conduct may facilitate meeting legal requirements.

10. Implementation schedule : Within six months from the issue of this policy, DST/DIT will bring out detailed implementation guidelines including the technology and standards for data and metadata. All Ministries/Departments will provide at least 5 high value datasets on data.gov.in within three months. Data.govt.in will only have the metadata and data itself will be accessed from the portals of the departments/ministries through the links from data portal. Existing mechanisms will review the implementation.

Metadata

a) Metadata documenting archived/online data sets of all types need to be made available when, or before, the dataset itself is released according to the terms above.

b) All metadata will follow standards and will minimally contain adequate information on proper citation, access, contact information, and discovery. Complete information including methods, structure, semantics, and quality control/assurance is expected for most datasets.

11. Responsibilities of Data Base Owners/ Data Generators/Controllers

a) The dataowners/generators/controllers (ministries/ departments/ organizations) shall:-

- extend authorization to database managers for access to information;
- authorize access to secondary users in written form;

b) Database managers shall:

- provide the day-to-day controls of the data;
- provide secondary users of how to interpret data;

c) The database owner shall validate data before the same is made accessible to the users.

d) The data owners, managers and all authorized secondary users shall take all reasonable precautions against unauthorized access, willful or not, to screens and/or reports containing sensitive data.

e) The Computing Services Department shall install security procedures to reasonably prevent unauthorized access to systems and data by students or any other unauthorized person.

12. Infrastructure provision

While policies provide official mandate, facilitation of optimum accessibility and usability of data by the implementers pre-suppose a trajectory of proper

organization of data, with access services and analysis tools that provide the researchers with added value. For data to be reused, it needs to be adequately described and linked to services that disseminate the data to other researchers and stakeholders. The current methods of storing data are as diverse as the disciplines that generate it. It is necessary to develop institutional repositories, data centres on domain and national levels that all methods of storing and sharing have to exist within the specific infrastructure to enable all users to access and use it.
