National Metadata Structure (NMDS)

A Standard Set of Guidelines for Ensuring Quality of Statistical Products Released

Objectives

- Promote reporting for each type of statistical process and its outputs across different Ministries/Divisions/Departments of NSO, hence facilitating comparisons across processes and outputs;

- Ensure that producer reports contain all the information required to facilitate identification of quality issues and potential improvements in statistical processes and their outputs; and

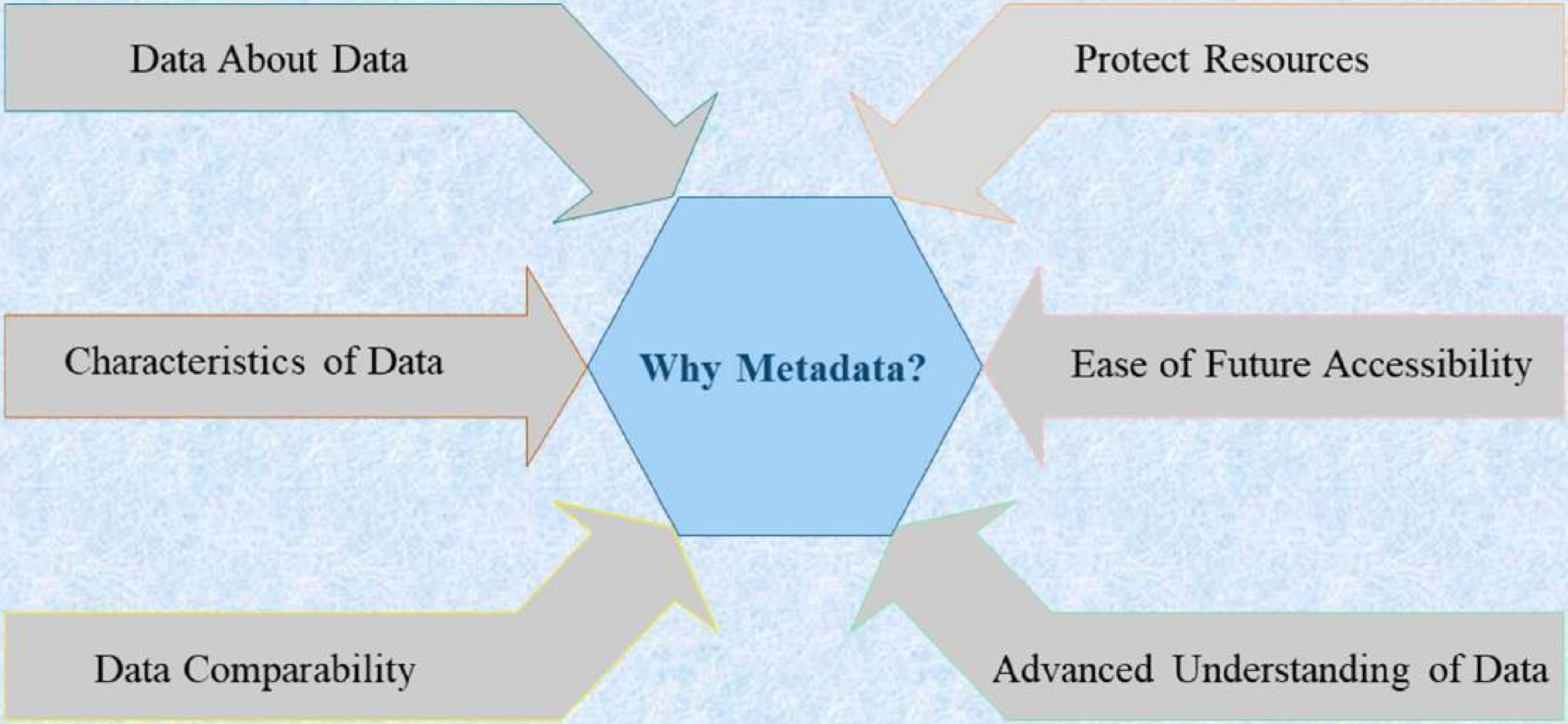
- Ensure that user reports contain all the information required by users to assess whether statistical outputs are fit for the purposes they have in mind.

What is metadata?

- It is structured information or documentation about data which:
  - informs users about the content, quality and condition of data;
  - describes the structure of datasets, explains, locates, or makes it easier to retrieve, use, or manage data;
  - provides information on the processes of data production;
  - guides on proper usage or interpretation of data.

- “Data that define and describe other data” (ISO definition)
- “Information about information” (Dion, 2006)

Why Metadata



Why Metadata?

Data About Data

Characteristics of Data

Data Comparability

Protect Resources

Ease of Future Accessibility

Advanced Understanding of Data

What does metadata do?

- Assist in retrieving and processing data
- Support correct use of data
- Provide transparency in data
- Enhance interoperability
- Improve archiving, preservation, institutional memory over time

Types of metadata

- **Structural metadata**
  - provide information about the structure of the dataset
  - act as identifiers and descriptors of the data, making it possible to properly identify, retrieve and browse the data
- **Reference metadata**
  - allow a thorough understanding and interpretation of the corresponding statistical data
  - Describe the concepts, definitions, methodology and quality of data; production and dissemination process, data access conditions, release policy, confidentiality, etc

Metadata standards (cont'd)

Two international metadata standards are becoming well established:

- SDMX (Statistical Data and Metadata Exchange)

  a number of international agencies have endorsed SDMX; supported by the UN Statistical Commission

- DDI (Data Dissemination Initiative)

  Microdata Management Toolkit of WB uses the DDI metadata standard

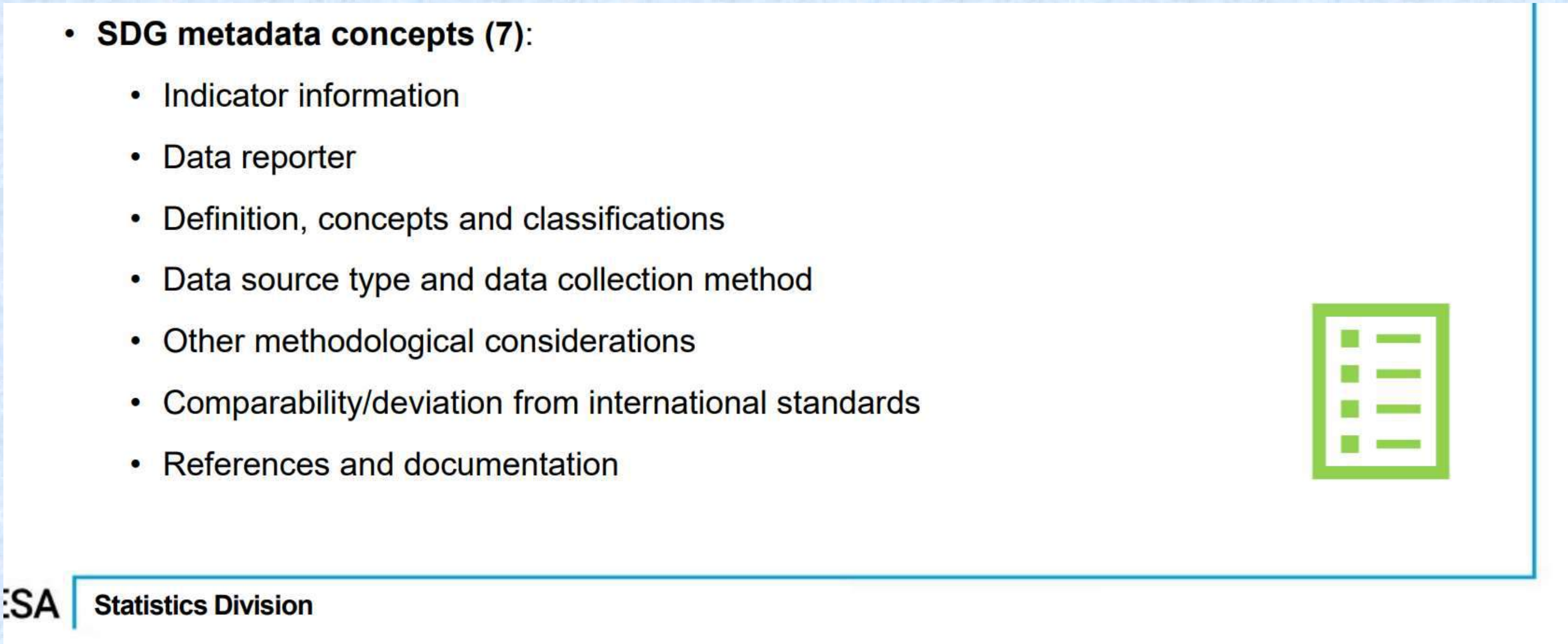
Users of metadata

Metadata support the knowledge of potential user of statistical information. The major users include:

- Users of statistical data
- Producers of statistical data
- Researchers on the development of statistical systems

Metadata Structure: International Organisations

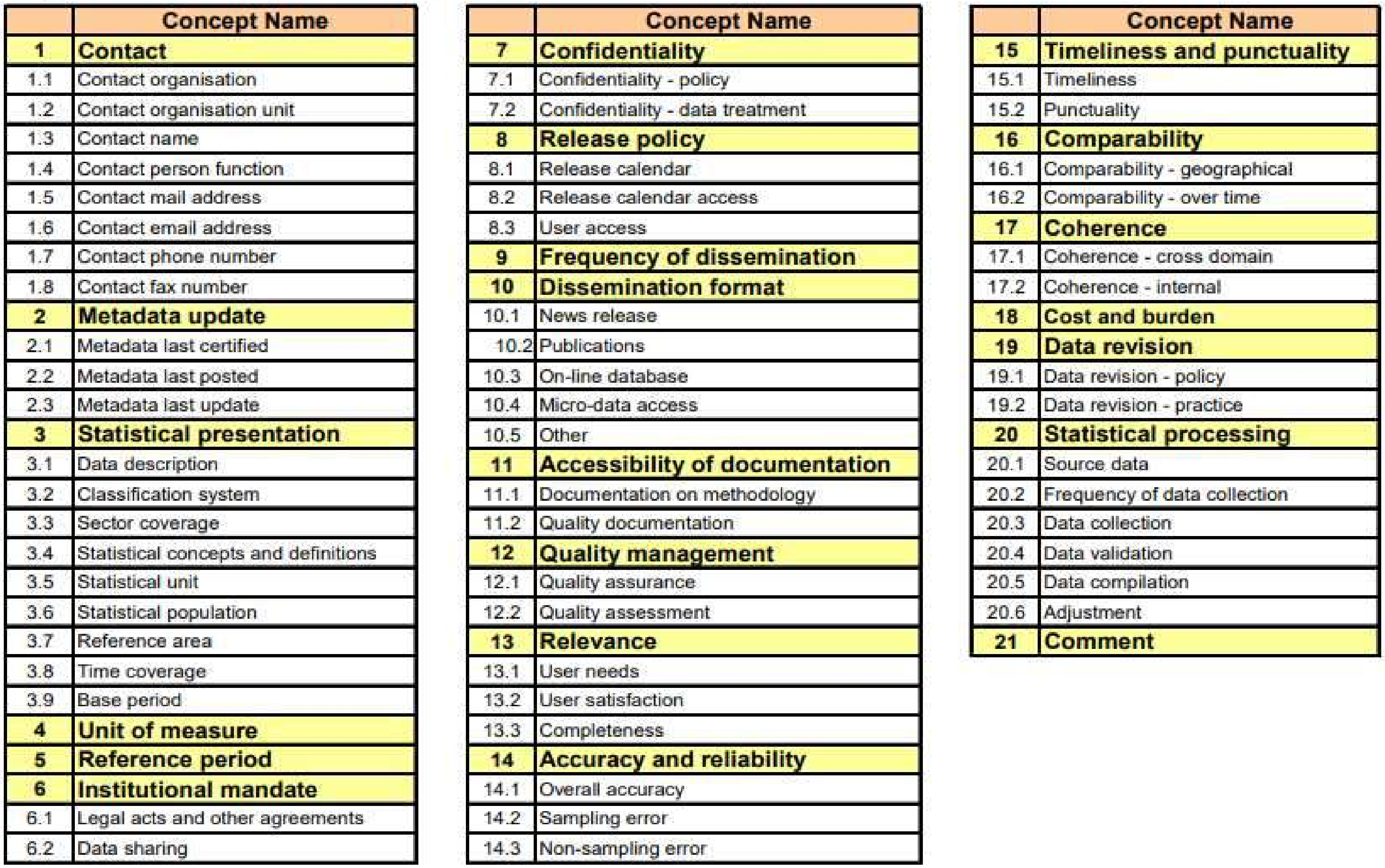
UN SDG Metadata Concept



- **SDG metadata concepts (7)**:
    - Indicator information
    - Data reporter
    - Definition, concepts and classifications
    - Data source type and data collection method
    - Other methodological considerations
    - Comparability/deviation from international standards
    - References and documentation

ESA | **Statistics Division**

Figure 2. The Euro SDMX Metadata Structure



	Concept Name
1	**Contact**
1.1	Contact organisation
1.2	Contact organisation unit
1.3	Contact name
1.4	Contact person function
1.5	Contact mail address
1.6	Contact email address
1.7	Contact phone number
1.8	Contact fax number
2	**Metadata update**
2.1	Metadata last certified
2.2	Metadata last posted
2.3	Metadata last update
3	**Statistical presentation**
3.1	Data description
3.2	Classification system
3.3	Sector coverage
3.4	Statistical concepts and definitions
3.5	Statistical unit
3.6	Statistical population
3.7	Reference area
3.8	Time coverage
3.9	Base period
4	**Unit of measure**
5	**Reference period**
6	**Institutional mandate**
6.1	Legal acts and other agreements
6.2	Data sharing
7	**Confidentiality**
7.1	Confidentiality - policy
7.2	Confidentiality - data treatment
8	**Release policy**
8.1	Release calendar
8.2	Release calendar access
8.3	User access
9	**Frequency of dissemination**
10	**Dissemination format**
10.1	News release
10.2	Publications
10.3	On-line database
10.4	Micro-data access
10.5	Other
11	**Accessibility of documentation**
11.1	Documentation on methodology
11.2	Quality documentation
12	**Quality management**
12.1	Quality assurance
12.2	Quality assessment
13	**Relevance**
13.1	User needs
13.2	User satisfaction
13.3	Completeness
14	**Accuracy and reliability**
14.1	Overall accuracy
14.2	Sampling error
14.3	Non-sampling error
15	**Timeliness and punctuality**
15.1	Timeliness
15.2	Punctuality
16	**Comparability**
16.1	Comparability - geographical
16.2	Comparability - over time
17	**Coherence**
17.1	Coherence - cross domain
17.2	Coherence - internal
18	**Cost and burden**
19	**Data revision**
19.1	Data revision - policy
19.2	Data revision - practice
20	**Statistical processing**
20.1	Source data
20.2	Frequency of data collection
20.3	Data collection
20.4	Data validation
20.5	Data compilation
20.6	Adjustment
21	**Comment**

MoSPI’s Metadata Structure

Brief Concepts of Metadata Structure

Item No	Concept name	Item No	Concept name	Item No	Concept name
1	**Contact**	3.2	Data sharing	7.2	Coherence
1.1	Contact Organisation	3.3	Release policy	**8**	**Statistical processing**
1.2	Compiling Agency	3.4	Release calendar	9.1	Source data type
1.3	Contact Details	3.5	Frequency of dissemination	9.2	Frequency of data collection
2	**Statistical presentation and description**	3.6	Data access	9.3	Data collection method
2.1	Data description	**4**	**Quality management**	9.4	Data validation
2.2	Classification system	4.1	Documentation on methodology	9.5	Data compilation
2.3	Sector coverage	4.2	Quality documentation	**9**	**Metadata update**
2.4	Statistical concepts and definitions	4.3	Quality assurance	9.1	Metadata last posted
2.5	Statistical unit	4.4	Quality assessment	9.2	Metadata last update
2.6	Statistical population	**5**	**Accuracy and reliability**		
2.7	Reference Period	5.1	Sampling error		
2.8	Base period	**6**	**Timeliness**		
2.9	Data Confidentiality	6.1	Timeliness		
3	**Institutional mandate**	**7**	**Coherence and comparability**		
3.1	Legal acts and other agreements	7.1	Comparability – over time		

Elements of Metadata Structure: Comparison amongst selected NSOs

Detailed Concepts of Metadata

Contact

Back to Brief Concepts

Item No	Concept name	Definition	Guidelines
1	**Contact**	**Individual or organisational contact points for the data or metadata, including information on how to reach the contact points.**	
1.1	Contact Organisation	The name of the organisation for data or metadata.	Provide the full name (not just acronym) of the organisation responsible for the processes and outputs (data and metadata) that are the subject of the report
1.2	Compiling Agency	Organisation collecting and/or elaborating the data being reported	Provide the full name of the Division under the organisation responsible for the processes and outputs (data and metadata) that are the subject of the report
1.3	Contact Details	The details of the contact points for the data or metadata.	Provide contact details of contact point(s) in following format: a. Organisation Name owning the processes and outputs: b. Author (if different from (a)) c. Disseminating Agency (if different from (a) and (b)) d. Name (first and last names): e. Designation: f. Postal address: g. email address (preferably designation based): h. Contact number: i. Fax number: If more than one name is provided, the details of main contact should be indicated. If the author of the report is different from the person(s) responsible for process and its outputs, provide this name also with his/her details

Statistical Presentation and Description

Back to Brief Concepts

2	**Statistical presentation and description**	**Description of the disseminated data which can be displayed to users as tables, graphs or maps**	
2.1	Data description	Main characteristics of the data set, referring to the data and indicators disseminated.	Describe briefly the main characteristics of the data in an easily and quickly understandable manner, referring to the main variables disseminated.
2.2	Classification system	Arrangement or division of objects into groups based on characteristics which the objects have in common	List all classifications and breakdowns that are used in the data (with their detailed names) and provide links (if publicly available). Type of dis-aggregation available in the data sets - for example rural-urban, male-female, etc. and whether data is available at the sub-national level or not, should be clearly specified.
2.3	Sector coverage	Main economic or other sectors	List the main economic or other sectors covered by the data and the size classes used, for example, Health/ Education/ Manufacturing etc
2.4	Statistical concepts and definitions	Statistical characteristics of statistical observations, variables	Define and describe briefly the main statistical variables that have been observed or derived. Indicate their types.
2.5	Statistical unit	Entity for which information is sought and for which statistics are ultimately compiled.	Define the type of statistical unit about which data are collected, e.g. enterprise, household, etc.
2.6	Statistical population	The total population of a defined class of people, objects or events	Define the target population of statistical units for which information is sought
2.7	Reference Period	The length of time for which data are available	State the time period(s) for which data is collected
2.8	Base period	The period of time used as the base of an index number, or to which a constant series refers.	Note that this concept applies only to certain types of outputs, such as indexes, for which a base period is defined and used. Eg, CPI, IIP, etc.
2.9	Data Confidentiality	Rules applied for treating the datasets to ensure statistical confidentiality and prevent unauthorised disclosure.	Describe the procedures that are used in protecting confidentiality.

Back to Brief Concepts

Institutional Mandate

3	**Institutional mandate**	**Law, set of rules or other formal set of instructions assigning responsibility as well as the authority to an organisation for the collection, processing, and dissemination of statistics**	
3.1	Legal acts and other agreements	Legal acts or other formal or informal agreements that assign responsibility as well as the authority to an agency for the collection, processing, and dissemination of statistics	State the national legal acts and/or other reporting agreements
3.2	Data sharing	Arrangements or procedures for data sharing and coordination between data producing agencies.	Describe the arrangements, procedures or agreements to facilitate data sharing and exchange between data producing agencies within the national statistical system
3.3	Release policy	Rules for disseminating statistical data to all interested parties	State if the release of the products is governed by some policy etc.
3.4	Release calendar	The schedule of statistical release dates.	State whether there is a release calendar for the statistical outputs from the process being reported, and if so, whether this calendar is publicly accessible and if yes, give a link or reference.
3.5	Frequency of dissemination	The time interval at which the statistics are disseminated over a given time period.	State the frequency with which the data are disseminated, e.g. monthly, quarterly, yearly.
3.6	Data access	The conditions and modalities by which users can access, use and interpret data	State the conditions and link on website from where the user can access the data For easy access of users, following details should also be mentioned about the dataset: **Title:** Name by which the data is known **Dataset Edition:** Edition of data (ex: first, second, final etc) **Dataset Reference data type:** Type of data entered in the field (ex: .txt, .dbf, .xls) **Presentation Format:** Presentation format of the data (ex: document, map, table, etc.) **Dataset Language:** language of any text in the data **Status/Version:** How updated is the data?

Quality Management

Back to Brief Concepts

4.	Quality management	Systems and frameworks in place within an organisation to manage the quality of statistical products and processes.	
4.1	Documentation on methodology	Descriptive text and references to methodological documents available.	List reference metadata files, methodological papers, summary documents and handbooks relevant to the statistical process.
4.2	Quality documentation	Documentation on procedures applied for quality management and quality assessment.	List relevant quality related documents, for example, other quality reports, studies.
4.3	Quality assurance	All systematic activities implemented that can be demonstrated to provide confidence that the processes will fulfil the requirements for the statistical output.	Describe the procedures (such as use of a general quality management system based on ISO 9000 series) to promote general quality management principles in the organisation. Describe the quality assurance framework used to implement statistical quality principles. Describe the quality assurance procedures specifically applied to the statistical process for which the report is being prepared, for example training courses, process monitoring, benchmarking, assessments, and use of best practices. Describe any ongoing or planned improvements in quality assurance procedures.
4.4	Quality assessment	Overall assessment of data quality, based on standard quality criteria.	Summarise the results of the most recent quality assessments and cross reference to the chapters in the report where the results are presented in more detail.

Back to Brief Concepts

Accuracy and Reliability

5	**Accuracy and reliability**	**Accuracy of data is the closeness of computations or estimates to the exact or true values that the statistics were intended to measure. Reliability of the data, defined as the closeness of the initial estimated value to the subsequent estimated value.**	
5.1	Sampling error	That part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a subset of the population is enumerated.	If probability sampling is used: • for user reports-provide the range of variation of the A1 indicator among key variables at user report level of detail; • for producer reports-provide the range of variation of the A1 indicator among key variables at producer report level of detail; • indicate the impact of sampling error on the overall accuracy of the results; • state how the calculation of sampling error is affected by adjustments for nonresponse, misclassifications and other sources of uncertainty, such as outlier treatment. If non-probability sampling is used: provide an assessment of representativeness, a motivation for the invoked model for estimation and risk of sampling bias

Timeliness and Punctuality

Back to Brief Concepts

6	Timeliness	The timeliness of the data collection release to be compiled.	
6.1	Timeliness	Length of time between data availability, the event or phenomenon the data describe, and final release to its users.	Outline the reasons for the time lag, if any. Outline efforts to reduce time lag in future.

Coherence and Comparability

Back to Brief Concepts

7	Coherence and comparability	Adequacy of statistics to be reliably combined in different ways and for various uses and the extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics	
7.1	Comparability – over time	The extent to which statistics are comparable or reconcilable over time	Provide information on possible limitations in the use of data for comparisons over time. Distinguish three broad possibilities: 1. There have been no changes, in which case this should be reported. 2. There have been some changes but not enough to warrant the designation of a break in series. 3. There have been sufficient changes to warrant the designation of a break in series.
7.6	Coherence	The extent to which statistics are reconcilable with System of National Accounts.	For producer reports only. Where relevant, the results of comparisons with the System of National Account framework / Other Statistical Standards and feedback from System of National Accounts / Other Statistical Standards with respect to coherence and accuracy problems should be reported and should be a trigger for further investigation.

Statistical Processing

← Back to Brief Concepts

8	Statistical processing	Any statistical processing undertaken to finalise the data	
8.1	Source data type	Characteristics and components of the raw statistical data used for compiling statistical aggregates.	Indicate if the data are based on a survey, administrative data, multiple data sources, or macro-aggregates. In the event of multisource or macro-aggregate processes describe each source. For each dataset from an administrative source, summarise the source, its primary purpose, and the most important data items acquired
8.2	Frequency of data collection	Time interval at which the source data are collected	Indicate the frequency of data collection (e.g. monthly, quarterly, annually, or continuous).
8.3	Data collection method	Method applied for gathering data for official statistics.	For each source of survey data: • describe the method(s) used to gather data from respondents; • annex or hyperlink the questionnaires(s). For each source of administrative data: • describe the acquisition process and how it was tested. For all sources: • describe the types of checks applied at the time of data entry.

Statistical Processing (Contd…)

Back to Brief Concepts

8.4	Data validation	Process of monitoring the results of data compilation and ensuring the quality of statistical results	Describe the procedures for checking and validating the source data and how the results are monitored and used. Describe the procedures for validating the aggregate output data (statistics) after compilation, including checking coverage and response rates, and comparing with data for previous cycles and with expectations. List other output datasets to which the data relate and outline the procedures for identifying inconsistencies between the output data and these other datasets
8.5	Data compilation	Operations performed on data to derive new information according to a given set of rules.	Describe the procedures for imputation, the most common reasons for imputation and imputation rates within each of the main strata. Describe the likely impact of imputation. Describe the procedures to derive new variables and to calculate aggregates and complex statistics. Describe the procedures for adjustment for non-response and the corrections to the design weights to account for differences in response rates. Describe the calculation of design weights, including calibration (if used). Describe the procedures for combining input data from different sources. Provide the ratio of the number of replaced values to the total number of values for a given variable. Specific reference to formula shall be made. The formula or mathematical equation used while computing different variables in the report may be described here in a structured format showing the Numerator; Denominator and Multiplier used for computing the same

Metadata Update

Back to Brief Concepts

9	Metadata update	The date on which the metadata element was inserted or modified in the database.	
9.1	Metadata last posted	Date of the latest dissemination of the metadata	The date when the complete set of metadata was last disseminated as a block should be provided (manually, or automatically by the metadata system).
9.2	Metadata last update	Date of last update of the content of the metadata.	The date when any metadata were last updated should be provided (manually, or automatically by the metadata system).

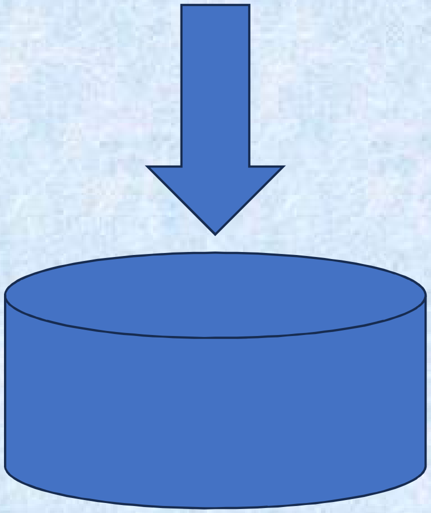
Metadata Structure for Statistical Products as Available in Public Domain

Use Cases of Metadata Compliance of MoSPI

What Next

- From ”paper” to database

- Information in the word document



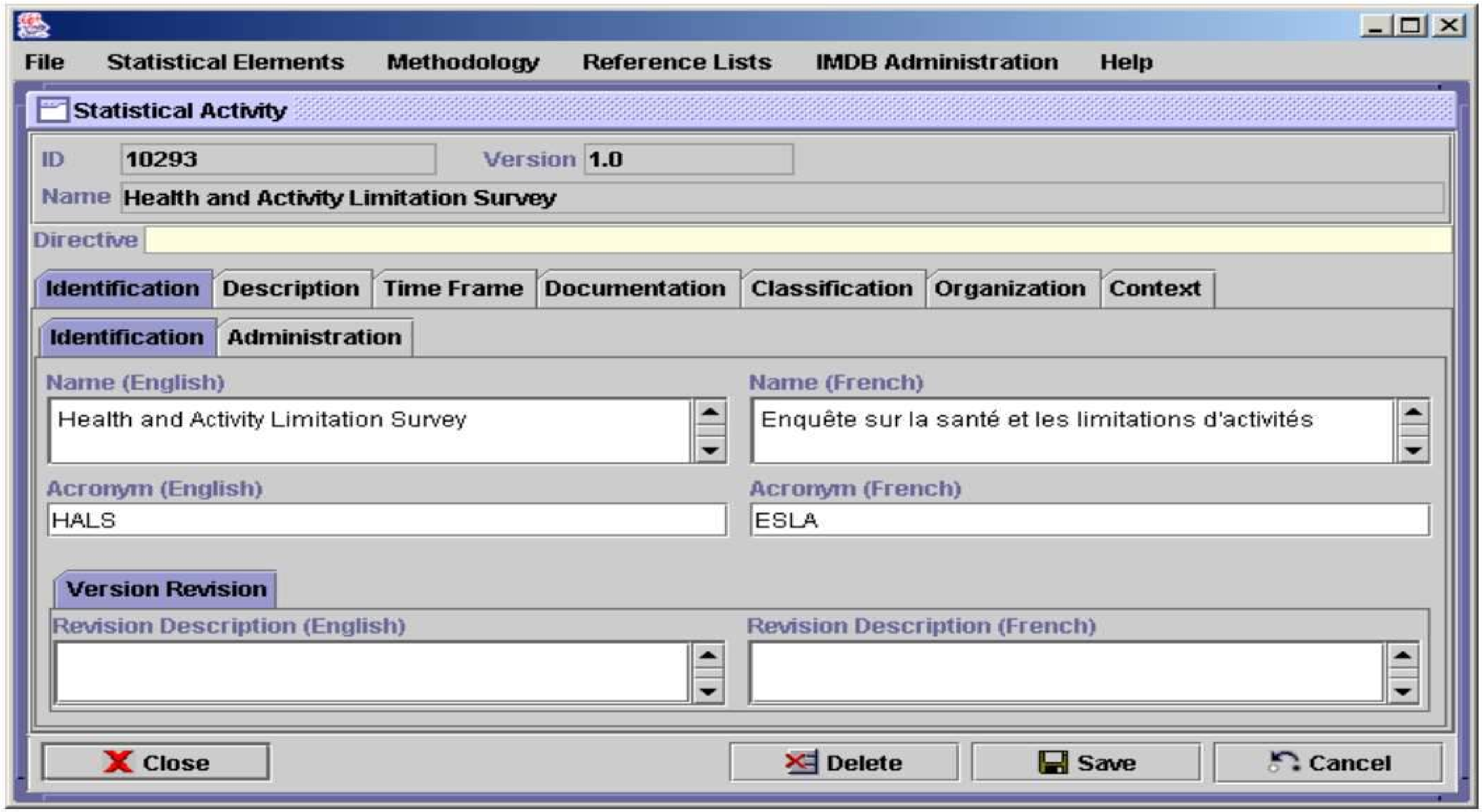
Revised NMDS

Introduction of Two new Concepts

- **2.3 International/National Standards Classification etc.**
- **8.6 Identifier of a statistical indicator**

NMDS

Loading metadata into India Integrated Metadata Base (IIMDB)


File
Statistical Elements
Methodology
Reference Lists
IMDB Administration
Help
Statistical Activity
Directive
Identification
Description
Time Frame
Documentation
Classification
Organization
Context
Identification
Administration
Version Revision
Close
Delete
Save
Cancel

ID 10293 Version 1.0

Name Health and Activity Limitation Survey

Name (English): Health and Activity Limitation Survey

Name (French): Enquête sur la santé et les limitations d'activités

Acronym (English): HALS

Acronym (French): ESLA

Revision Description (English):

Revision Description (French):

Thank You

सत्यमेव जयते

Government of India

Ministry of Statistics & Programme Implementation

DATA DISSEMINATION :

NATIONAL METADATA STRUCTURE (NMDS) FOR STATISTICAL PRODUCTS

METADATA
author
title
subtitle
pen name
...
description
keywords
series

Government of India
Ministry of Statistics and Programme Implementation
Policy Implementation & Monitoring Division (PIMD)
Khurshid Lal Bhawan, 6th Floor
Janpath, New Delhi – 110001

☎ +91-11-23455610 pimd-coord@mospi.gov.in

PIMD/Jul/2024/2

Government of India

Ministry of Statistics and Programme Implementation

Data Dissemination: National Metadata Structure (NMDS) For Statistical Products

July 2024

Government of India
Ministry of Statistics and Programme Implementation
Policy Implementation & Monitoring Division (PIMD)
Khurshid Lal Bhawan, 6th Floor
Janpath, New Delhi – 110001
Ph: (011) 23455610

Data Dissemination: National Metadata Structure (NMDS) For Statistical Products (Revised)

July 2024

Foreword

Introduction

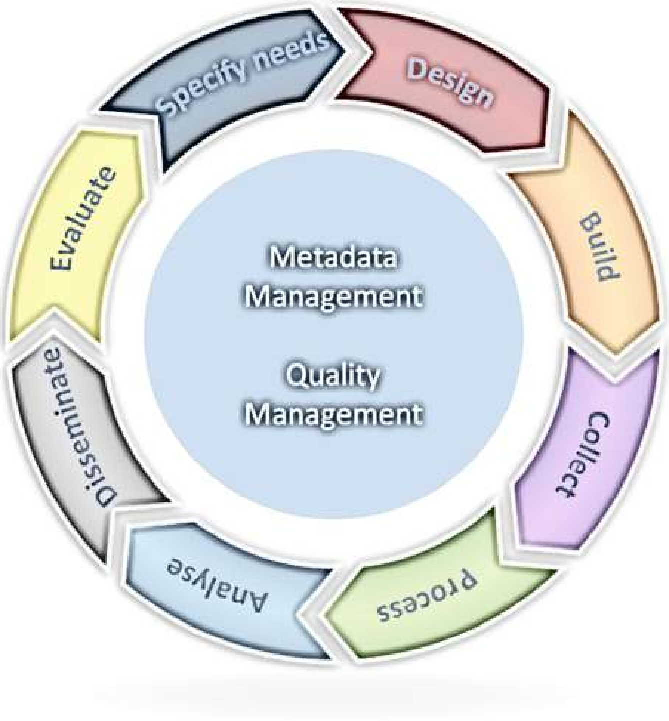
National Statistical Office (NSO), Ministry of Statistics & Programme Implementation, presents and disseminates data and metadata through different products like Census data (Economic Census); Survey data such as NSS Surveys viz. Household Surveys, etc, Annual Survey of Industries (ASI), Consumer Price Indices (CPI), and macro-economic aggregates like National Income, Index of Industrial Production (IIP). In addition, statistical data is presented in analytical publications such as NSS Reports, Annual Survey of Industries Reports, National Indicator Framework (NIF) for monitoring SDGs, Energy Statistics, EnviStats India, Women & Men in India etc., which provide analysis of data, supported by the visual presentation of that data in the form of graphs and maps.

The production of data and presentation of metadata structure requires an overview of the arrangements, technical infrastructure and skills required for a holistic and integrated approach to the presentation and dissemination of statistical data and metadata to different user groups. National Metadata Structure (NMDS) is to provide guidelines for the data producer to adhere to a basic minimum quality standard in order to establish and maintain the quality of data and enhance ease in sharing data. The specific objectives of this document are:

- to promote reporting for each type of statistical process and its outputs across different Ministries/Divisions/Departments of NSO, hence facilitating comparisons across processes and outputs;
- to ensure that producer reports contain all the information required to facilitate identification of quality issues and potential improvements in statistical processes and their outputs; and
- to ensure that user reports contain all the information required by users to assess whether statistical outputs are fit for the purposes they have in mind.

A. What is Metadata?

A.1. Metadata should contain all the information users need to analyse a dataset and draw conclusions. It increases data accessibility by summarizing the most important information (i.e. methodology, sampling design, interview mode, etc.) required for analyzing a dataset which alleviates the need for users to search for supporting documents and reports. Furthermore, good metadata clearly articulates the potential uses for a dataset, preventing potential misuses. Metadata is also a tool for rendering complex microdata structures into something meaningful, navigable, and user-friendly. Finally, the adoption of well-known metadata schemas and vocabularies allows for semantic interoperability.


Specify needs
Design
Build
Collect
Process
Analyse
Disseminate
Evaluate
Metadata Management
Quality Management

The Metadata process is fully integrated in the Generic Statistical Business Process Model[1] (GSBPM) which has metadata as one of the key elements in the version 5.1.

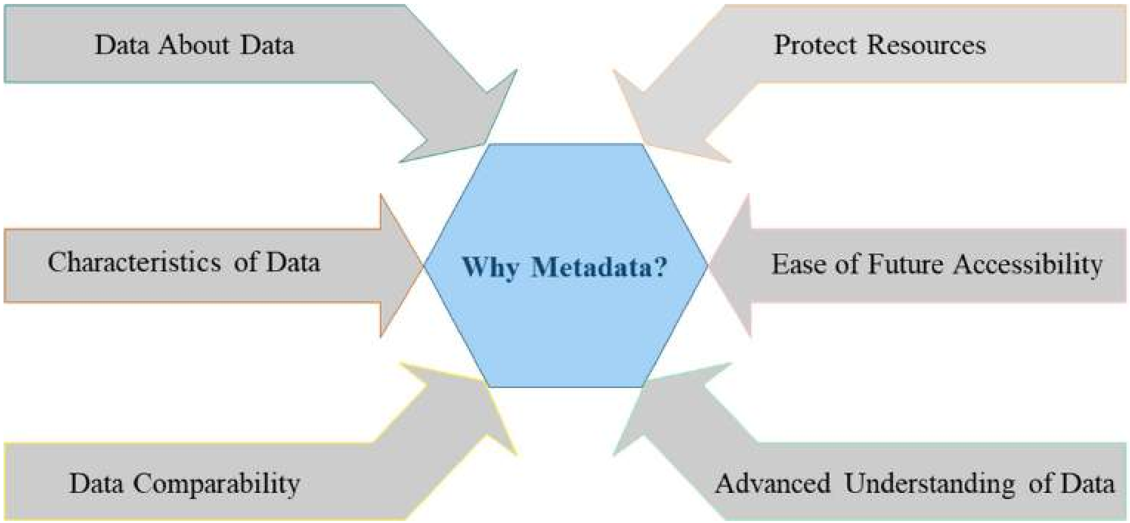
---

[1] UNECE: United Nations Economic Commission for Europe, https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1

B. Why Metadata?

B.1. In most information technology usages, the prefix of meta conveys "an underlying definition or description." So it is that, at its most basic, metadata is data about data. More precisely, however, metadata describes data containing specific information like type, length, textual description and other characteristics. Metadata makes it much easier to find relevant data and to use a dataset, users need to understand how the data is structured, definitions of terms used, how it was collected, and how it should be read.

B.2. Metadata is an important way to protect resources and their future accessibility. For archiving and preservation purposes, it takes metadata elements that track the object's lineage, and describe its physical characteristics and behaviour so it can be replicated on technologies in the future.


Data About Data
Protect Resources
Characteristics of Data
Why Metadata?
Ease of Future Accessibility
Data Comparability
Advanced Understanding of Data

B.3. In today's modern data driven world and in the era of digital transactions, huge amount of data is generated on real time basis, and lately, a large number of organisations/agencies have started producing data, the quantum of which is huge, and thus arises a need of standard regulatory framework to be laid down to assure the quality of data produced by different producers. It will also serve the purpose of ensuring data comparability across time horizons so as to enable better understanding of different social and economic movements.

B.4. Although metadata may not seem exciting or impressive, the true importance of metadata can never be underrated and hence, is important to take a concerted effort to build sound metadata structure to draw maximum gains from varied data sets.

C. Role of MoSPI in Building Metadata Structure

C.1. MoSPI being a nodal agency for planned development of the statistical system in the country is also responsible for maintaining the highest standards of data quality which adhere to basic guidelines of International Agencies so as to ensure India’s statistical system is one of the frontrunners in quality data producer. MoSPI aims at raising the National Statistical System (NSS) to the epitome of being one of the best professionally equipped government data producing agencies by building the best of IT infrastructure among others in the system, and Metadata is one of the building blocks to achieve the objective.

C.2. The document presents the NMDS in two formats- the first one is the indexed version as NMDS concepts (Section F), and the second one presents details of concepts through definition and guidelines (Section G).

D. Metadata Management

D.1 It is advisable to put in place a metadata policy by the official statistical producing agencies, ab initio. The policy is a set of broad, high level principles that form the guiding framework within which metadata management can operate.

D.2 Once the metadata policy is put in place, for an organisation, metadata should be compiled and maintained actively. Otherwise, the currency, and thus use of Metadata will degrade with time. To realise the full capabilities of Metadata, it is necessary that the Metadata are maintained over a long period of time. Even with investment in technically sophisticated search tools, such systems may find little stakeholders acceptance, if the data are incomplete or is not updated regularly.

While preparing the NMDS, the following core principles should be borne in mind:

i. Metadata Handling:
  a. Statistical Business Process Model
  b. Active, not Passive
  c. Reuse for Efficiency
  d. Version Preservations

ii. Metadata Authority
  e. Registration
  f. Single Source
  g. One Entry/Update
  h. Standards Variations

iii. Relationship to Statistical Business Processes
  i. Integrity
  j. Matching Metadata
  k. Describe Flow
  l. Capture at Source
  m. Exchange and Use

iv. Users
  n. Identify Users
  o. Variant Formats
  p. Availability

E. Retention, Preservation, and Destruction

National Statistics constitute valuable and irreplaceable assets whose value can increase through widespread and long-term use. National Statistics should thus be backed by the Data Management Policy eliciting the arrangements it has in place for the retention, long term preservation, and destruction of its resources including Metadata.

F. National Metadata Structure (NMDS) Concepts - Index

Item No	Concept name
1	**Contact**
1.1	Contact Organisation
1.2	Compiling Agency
1.3	Contact Details
2	**Statistical Presentation and Description**
2.1	Data description
2.2	Classification system
2.3	International/National Standards Classification etc.
2.4	Sector coverage
2.5	Statistical concepts and definitions
2.6	Statistical unit
2.7	Statistical population
2.8	Reference Period
2.9	Data Confidentiality
3	**Institutional Mandate**
3.1	Legal acts and other agreements
3.2	Data sharing
3.3	Release policy
3.4	Release calendar
3.5	Frequency of dissemination
3.6	Data access
4	**Quality Management**
4.1	Documentation on methodology
4.2	Quality documentation
4.3	Quality assurance
4.4	Quality assessment
5	**Accuracy and Reliability**
5.1	Sampling error
6	**Timeliness**
6.1	Timeliness
7	**Coherence and Comparability**
7.1	Comparability – over time
7.2	Coherence
8	**Statistical Processing**
8.1	Source data type
8.2	Frequency of data collection
8.3	Data collection method
8.4	Data validation
8.5	Data compilation
8.6	Identifier of a statistical indicator
9	**Metadata Update**
9.1	Metadata last posted
9.2	Metadata last update

G. Details of NMDS Concepts

Item No	Concept name	Definition	Guidelines
1	**Contact**	**Individual or organisational contact points for the data or metadata, including information on how to reach the contact points.**	
1.1	Contact Organisation	The name of the organisation of the contact points for the data or Metadata.	Provide the full name (not just acronym/code name) of the organisation responsible for the processes and outputs (data and metadata) that are the subject of the report
1.2	Compiling agency	Organisation collecting and/or elaborating the data being reported	Provide the full name of the Department/Division under the organisation responsible for the processes and outputs (data and metadata) that are the subject of the report
1.3	Contact Details	The details of the contact points for the data or metadata.	Provide contact details of contact point(s) in following format: a. Name of Organisation owning the processes and outputs b. Author (if different from (a)) c. Disseminating Agency (if different from (a) and (b)) d. Name (first and last names) e. Designation f. Postal address g. email address (preferably designation based) h. Contact number i. Fax number If more than one name is provided, the details of main contact should be indicated. If the author of the report is different from the person(s) responsible for process and its outputs, provide this name also with his/her details
2	**Statistical Presentation and Description**	**Description of the disseminated data which can be displayed to users as tables, graphs or maps**	

Item No	Concept name	Definition	Guidelines
2.1	Data description	Main characteristics of the data set, referring to the data and indicators disseminated.	Describe briefly the main characteristics of the data in an easily and quickly understandable manner, referring to the main variables disseminated.
2.2	Classification system	Arrangement or division of objects into groups based on characteristics which the objects have in common	List all classifications and breakdowns that are used in the data (with their detailed names) and provide links (if publicly available). Type of dis-aggregation available in the data sets - for example rural-urban, male-female, etc. and whether data is available at the sub-national level or not, should be clearly specified.
2.3	Sector coverage	Main economic or other sectors	List the main economic or other sectors covered by the data and the size classes used, for example, Health/ Education/ Manufacturing etc for sectors and classes based on number of employees for size classes.
2.4	International/ National Standards Classification etc.	International/ National standard classification, is the primary tool for collecting and presenting internationally comparable statistics by economic activity.	Mention the International/ National Standard classification used.
2.5	Statistical concepts and definitions	Statistical characteristics of statistical observations, variables	Define and describe briefly the main statistical variables that have been observed or derived. Indicate their types.
2.6	Statistical unit	Entity for which information is sought and for which statistics are ultimately compiled.	Define the type of statistical unit about which data are collected, e.g. enterprise, household, etc.
2.7	Statistical population	The total population of a defined class of people, objects or events	Define the target population of statistical units for which information is sought. For example, agricultural household, general household, industrial unit, etc.

Item No	Concept name	Definition	Guidelines
2.8	Reference Period	The length of time for which data are available	State the time period(s) for which data is collected
2.9	Data Confidentiality[2]	Rules applied for treating the datasets to ensure statistical confidentiality and prevent unauthorised disclosure.	Describe the procedures that are used in protecting confidentiality, viz., anonymisation, legal provision, if any.
3	**Institutional Mandate**	**Law, set of rules or other formal set of instructions assigning responsibility as well as the authority to an organisation for the collection, processing, and dissemination of statistics**	
3.1	Legal acts and other agreements	Legal acts or other formal or informal agreements that assign responsibility as well as the authority to an agency for the collection, processing, and dissemination of statistics	State the national legal acts and/or other reporting agreements
3.2	Data sharing	Arrangements or procedures for data sharing and coordination between data producing agencies.	Describe the arrangements, procedures or agreements to facilitate data sharing and exchange between data producing agencies within the national statistical system
3.3	Release policy	Rules for disseminating statistical data to all interested parties	State if the release of the products is governed by some policy etc.
3.4	Release calendar	The schedule of statistical release dates.	State whether there is a release calendar for the statistical outputs from the process being reported, and if so, whether this calendar is publicly accessible and if yes, give a link or reference.

---

[2] All statistical information published by any agency shall be arranged in such a manner so as to prevent any particulars becoming identifiable by any person (other than the informant by whom those particulars were supplied) as the particulars relating to the informant who supplied it, even though the process of elimination (Source: Collection of Statistics Act, 2008).

Item No	Concept name	Definition	Guidelines
3.5	Frequency of dissemination	The time interval at which the statistics are disseminated over a given time period.	State the frequency with which the data are disseminated, e.g. monthly, quarterly, yearly.
3.6	Data access	The conditions and modalities by which users can access, use and interpret data	State the conditions and link on website from where the user can access the data For easy access of users, following details should also be mentioned about the dataset: **Title:** Name by which the data is known **Dataset Edition:** Edition of data (ex: first, second, final etc) **Dataset Reference data type:** Type of data entered in the field (ex: .txt, .dbf, .xls) **Presentation Format:** Presentation format of the data (ex: document, map, table, etc.) **Dataset Language:** language of any text in the data **Status/Version:** How updated is the data?
4	**Quality Management**	**Systems and frameworks in place within an organisation to manage the quality of statistical products and processes.**	
4.1	Documentation on methodology	Descriptive text and references to methodological documents available.	List reference metadata files, methodological papers, summary documents and handbooks relevant to the statistical process.
4.2	Quality documentation	Documentation on procedures applied for quality management and quality assessment.	List relevant quality related documents, for example, other quality reports, studies.

Item No	**Concept name**	**Definition**	**Guidelines**
4.3	Quality assurance	All systematic activities implemented that can be demonstrated to provide confidence that the processes will fulfil the requirements for the statistical output.	Describe the procedures (such as use of a general quality management system based on ISO 9000 series) to promote general quality management principles in the organisation. Describe the quality assurance framework used to implement statistical quality principles. Describe the quality assurance procedures specifically applied to the statistical process for which the report is being prepared, for example training courses, process monitoring, benchmarking, assessments, and use of best practices. Describe any ongoing or planned improvements in quality assurance procedures.
4.4	Quality assessment	Overall assessment of data quality, based on standard quality criteria.	Summarise the results of the most recent quality assessments and cross reference to the chapters in the report where the results are presented in more detail.
5	**Accuracy and Reliability**	**Accuracy of data is the closeness of computations or estimates to the exact or true values that the statistics were intended to measure. Reliability of the data, defined as the closeness of the initial estimated value to the subsequent estimated value.**	
5.1	Sampling error	That part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a subset of the population is enumerated.	If probability sampling is used: • for user reports-provide the range of variation of the A1[3] indicator among key variables at user report level of detail; • for producer reports-provide the range of variation of the A1 indicator among key variables at producer report level of detail;

---

[3] In international terminology, A1 is simply used for Sampling Errors. In other words, indicator for Sampling Errors is denoted by A1. For further details, refer to: https://ec.europa.eu/eurostat/documents/64157/4373903/02-ESS-Quality-and-performance-Indicators-2014.pdf/5c996003-b770-4a7c-9c2f-bf733e6b1f31

Item No	**Concept name**	**Definition**	**Guidelines**
			• indicate the impact of sampling error on the overall accuracy of the results; • state how the calculation of sampling error is affected by adjustments for nonresponse, misclassifications and other sources of uncertainty, such as outlier treatment. If non-probability sampling is used: provide an assessment of representativeness, a motivation for the invoked model for estimation and risk of sampling bias
6	**Timeliness**	**The timeliness of the data collection release to be compiled.**	
6.1	Timeliness	Length of time between data availability, the event or phenomenon the data describe, and final release to its users.	Outline the reasons for the time lag, if any. Outline efforts to reduce time lag in future.
7	**Coherence and Comparability**	**Adequacy of statistics to be reliably combined in different ways and for various uses and the extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics**	
7.1	Comparability – over time	The extent to which statistics are comparable or reconcilable over time	Provide information on possible limitations in the use of data for comparisons over time. Distinguish three broad possibilities: 1. There have been no changes, in which case this should be reported. 2. There have been some changes but not enough to warrant the designation of a break in series. 3. There have been sufficient changes to warrant the designation of a break in series.

Item No	Concept name	Definition	Guidelines
7.2	Coherence	The extent to which statistics are reconcilable with System of National Accounts	For producer reports only. Where relevant, the results of comparisons with the System of National Account framework / Other Statistical Standards and feedback from System of National Accounts / Other Statistical Standards with respect to coherence and accuracy problems should be reported and should be a trigger for further investigation.
8	**Statistical Processing**	**Any statistical processing undertaken to finalise the data**	
8.1	Source data type	Characteristics and components of the raw statistical data used for compiling statistical aggregates.	Indicate if the data are based on a survey, administrative data, multiple data sources, or macro-aggregates. In the event of multisource or macro-aggregate processes describe each source. For each dataset from an administrative source, summarise the source, its primary purpose, and the most important data items acquired
8.2	Frequency of data collection	Time interval at which the source data are collected	Indicate the frequency of data collection (e.g. monthly, quarterly, annually, or continuous).
8.3	Data collection method	Method applied for gathering data for official statistics	For each source of survey data: • describe the method(s) used to gather data from respondents; • annex or hyperlink the questionnaires(s). For each source of administrative data: • describe the acquisition process and how it was tested. For all sources: • describe the types of checks applied at the time of data entry.

Item No	Concept name	Definition	Guidelines
8.4	Data validation	Process of monitoring the results of data compilation and ensuring the quality of statistical results	Describe the procedures for checking and validating the source data and how the results are monitored and used. Describe the procedures for validating the aggregate output data (statistics) after compilation, including checking coverage and response rates, and comparing with data for previous cycles and with expectations. List other output datasets to which the data relate and outline the procedures for identifying inconsistencies between the output data and these other datasets
8.5	Data compilation	Operations performed on data to derive new information according to a given set of rules.	Describe the procedures for imputation, the most common reasons for imputation and imputation rates within each of the main strata. Describe the likely impact of imputation. Describe the procedures to derive new variables and to calculate aggregates and complex statistics. Describe the procedures for adjustment for non-response and the corrections to the design weights to account for differences in response rates. Describe the calculation of design weights, including calibration (if used). Describe the procedures for combining input data from different sources. Provide the ratio of the number of replaced values to the total number of values for a given variable. Specific reference to formula shall be made. The formula or mathematical equation used while computing different

Item No	**Concept name**	**Definition**	**Guidelines**
			variables in the report may be described here in a structured format showing the Numerator; Denominator and Multiplier used for computing the same
8.6	Identifier of a statistical indicator	A label used to identify a statistical indicator. It may have a form, a short alphanumeric name or a code	Mention the name of statistical indicator. Describe the attribute of a statistical indicator. Mention the value of a statistical indicator.
9	**Metadata Update**	**The date on which the metadata element was inserted or modified in the database.**	
9.1	Metadata last posted	Date of the latest dissemination of the metadata	The date when the complete set of metadata was last disseminated as a block should be provided (manually, or automatically by the metadata system).
9.2	Metadata last update	Date of last update of the content of the metadata.	The date when any metadata were last updated should be provided (manually, or automatically by the metadata system).

Concept name
Contact Organisation
Compiling agency
Contact Details
Data description
Classification system
Sector coverage
Statistical concepts and definitions
Statistical unit
Statistical population
Reference Period
Data Confidentiality
Legal acts and other agreements
Data sharing
Release policy
Release calendar
Frequency of dissemination
Data access
Documentation on methodology
Quality documentation
Quality assurance
Quality assessment
Sampling error
Timeliness
Comparability – over time

Coherence – System of National Accounts
Source data type
Frequency of data collection
Data collection method
Data validation
Data compilation
Metadata last posted
Metadata last update

Definition	India
The name of the organisation of the contact points for the data or metadata.	Yes
Organisation collecting and/or elaborating the data being reported	Yes
The details of the contact points for the data or metadata.	Yes
Main characteristics of the data set, referring to the data and indicators disseminated.	Yes
Arrangement or division of objects into groups based on characteristics which the objects have in common	Yes
Main economic or other sectors	Yes
Statistical characteristics of statistical observations, variables	Yes
Entity for which information is sought and for which statistics are ultimately compiled.	Yes
The total population of a defined class of people, objects or events	Yes
The length of time for which data are available	Yes
Rules applied for treating the datasets to ensure statistical confidentiality and prevent unauthorised disclosure.	Yes
Legal acts or other formal or informal agreements that assign responsibility as well as the authority to an agency for the collection, processing, and dissemination of statistics	Yes
Arrangements or procedures for data sharing and coordination between data producing agencies.	Yes
Rules for disseminating statistical data to all interested parties	Yes
The schedule of statistical release dates.	Yes
The time interval at which the statistics are disseminated over a given time period.	Yes
The conditions and modalities by which users can access, use and interpret data	Yes
Descriptive text and references to methodological documents available.	Yes
Documentation on procedures applied for quality management and quality assessment.	Yes
All systematic activities implemented that can be demonstrated to provide confidence that the processes will fulfil the requirements for the statistical output.	Yes
Overall assessment of data quality, based on standard quality criteria.	Yes
That part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a subset of the population is enumerated.	Yes
Length of time between data availability, the event or phenomenon the data describe, and final release to its users.	Yes
The extent to which statistics are comparable or reconcilable over time	Yes

The extent to which statistics are reconcilable with System of National Accounts / Other Statistical Standards	Yes
Characteristics and components of the raw statistical data used for compiling statistical aggregates.	Yes
Time interval at which the source data are collected	Yes
Method applied for gathering data for official statistics	Yes
Process of monitoring the results of data compilation and ensuring the quality of statistical results	Yes
Operations performed on data to derive new information according to a given set of rules.	Yes
Date of the latest dissemination of the metadata	Yes
Date of last update of the content of the metadata.	Yes

UK	Canada	NZ	Australia	Sweden
Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes
No	Yes	Yes	No	Yes
Yes	Yes	Yes	Yes	Yes
Yes	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	Yes	Yes	Yes
No	Yes	No	No	No
No	No	No	No	Yes
No	No	No	Yes	No
No	No	No	No	Yes
No	No	No	No	No
No	No	No	No	Yes
Yes	No	Yes	Yes	Yes
Yes	No	No	Yes	No
No	No	No	No	No
No	Yes	No	No	Yes
No	Yes	No	No	Yes
No	Yes	No	No	Yes
No	No	No	No	No
No	No	No	No	Yes
No	No	No	No	No

No	No	No	No	No
Yes	Yes	No	Yes	Yes
Yes	No	No	Yes	No
Yes	No	Yes	Yes	Yes
No	No	No	No	No
No	No	No	No	No
Yes	No	Yes	Yes	Yes
Yes	No	Yes	Yes	Yes

Categories
Surveys and statistical programs
Questionnaires
Variables and statistical units
Statistical classifications
Public consultation and notices
Methodology products and services
Metadata Frameworks
International links

https://www.statcan.gc.ca/eng/concepts/index

Details
Alphabetical list
List by subject
Alphabetical list
List by subject
Variables
Statistical units
Gender and sex variables
Chart of Accounts: Financial position and performance of a business
Environmental classifications
Geographic classifications
Industry classifications
Institutional units and sectors
Instructional programs
Occupational classifications
Product classifications
Classifications - A to Z
Classifying Cannabis
Classification search tools
Concordances between classifications
Variables and statistical units
Geographic classifications
Industry classifications
Instructional program classifications
Occupational classifications
Product classifications
Research and development classifications
Consulting services in statistical and survey methods
Data Quality Toolkit (12-606-X)
Dissemination Geography Unique Identifier
Standard table symbols
Statistics Canada Quality Guidelines (12-539-X)
Statistics Canada's Quality Assurance Framework (12-586-X)
Survey methodology journal
Survey methods and practices manual
Generic Statistical Information Model (GSIM)
Generic Statistical Business Process Model (GSBPM)
Record Linkage Project Process Model (12-605-X)
Trilateral website for NAICS and NAPCS
United Nations Economic Commission for Europe (UNECE)
United Nations Statistics Division (UNSD)
Voorburg group on services statistics

Label
Series
Classifications
Concepts
Codes
Populations
Questionnaires and forms
Variables

http://datainfoplus.stats.govt.nz/?_g

Description
Search or browse a list of series we produce. Find descriptions of series, including their frequen
Search a list of classifications we use.
Search or browse for information about the concepts explored in our studies, including the statis
Search for classification codes by keyword or browse for codes by classification.
Search or browse a list of populations we study.
Search a list of the instruments used in our studies.
Search or browse a list of variables used in our studies, including their descriptions.

a=2.123204040.271814164.1627020859-1370231482.1627020859 cy and collection methods.

stical terms we use.

Label
creator
date
Date Available
Date Created
Date Copyrighted
Date Licensed
Date Issued
Date Modified
Date Valid
title
Alternative Title
availability
identifier
Bibliographic Citation
publisher
description
function
language
subject
type
Aggregation Level
Type Category
Document Type
Service Type
audience
contributor
coverage
Australian Jurisdiction
Spatial Coverage
Temporal Coverage
format
Extent
Medium
mandate
Act
Case
Regulation
relation
Conforms To
Has Format
Has Part

Has Version
Is Basis For
Is Based On
Is Format Of
Is Part Of
Is Referenced By
Is Replaced By
Is Required By
Is Version Of
Replaces
Requires
rights
Access Rights
License
Protective Marking
Rights Holder
source
corporateName
Country
Electronic mail address
Facsimile number
Locality name
Personal Name
Physical Address
Position Name
Postal Address
Postcode
Role
Sector
State or Territory
Telephone number
Website location
corporateName
Cost
Country
email
fax
Hours of service
instructions
Locality name
Personal Name
physicalAccess
Physical Address

Position Name
Postal Address
Postcode
Role
Sector
State or Territory
Telephone number
Website location
Metadata Date Stamp
Metadata File Identifier
Metadata Character Set
Metadata Contact
Metadata Language
Metadata Standard Name
Metadata Standard Version
Metadata Update Date

Definition
An entity primarily responsible for making the resource.
A point or period of time associated with an event in the life of the resource
Date (often a range) that the resource became or will be available.
Date of creation of the resource
Date of copyright of the resource.
Date a license was applied or became effective
Date of formal issuance (e.g. publication) of the resource
Date on which the resource was changed.
Date (often a range) of validity of a resource.
A name given to the resource.
An alternative name for the resource.
How the resource can be obtained or accessed, or contact information for obtaining the resource
An unambiguous reference to the resource within a given context.
A bibliographic reference for the resource.
An entity responsible for making the resource available.
An account of the resource.
The business function to which the resource relates.
A language of the resource.
The topic of the resource.
The nature or genre of the resource.
The level of aggregation of the described resource
The generic type of the resource being described
The form of the described resource where the value of category is 'document'.
The form of the described resource where the value of category is 'service'.
A class of entity for whom the resource is intended or useful.
An entity responsible for making contributions to the resource.
The spatial or temporal topic of the resource, the spatial applicability of the resource, or the juris
The name of the political/administrative entity covered by the described resource.
Spatial characteristics of the resource.
Temporal characteristics of the resource.
The file format, physical medium, or dimensions of the resource.
The size or duration of the resource.
The material or physical carrier of the resource.
A specific legal instrument which requires or drives the creation or provision of the resource.
A specific piece of legislation which requires or drives the creation or provision of the resource.
A specific piece of case law which requires or drives the creation or provision of the resource.
A specific regulation which requires or drives the creation or provision of the resource.
A related resource.
An established standard to which the described resource conforms.
A related resource that is substantially the same as the pre-existing described resource, but in an
A related resource that is included either physically or logically in the described resource.

A related resource that is a version, edition or adaptation of the described resource.
A related resource that is a performance, production, derivation, translation or interpretation of
A related resource of which the described resource is a performance, production, derivation, tra
A related resource that is substantially the same as the described resource, but in another forma
A related resource in which the described resource is physically or logically included.
A related resource that references, cites or otherwise points to the described resource.
A related resource that supplants, displaces or supersedes the described resource.
A related resource that requires the described resource to support its function, delivery or coher
A related resource of which the described resource is a version, edition or adaptation.
A related resource that is supplanted, displaced or superseded by the described resource.
A related resource that is required by the described resource to support its function, delivery or
Information about rights held in and over the resource.
Information about who can access the resource.
A legal document giving official permission to do something withthe resource.
A protective marking applied to the described resource.
A person or organisation owning or managing rights over the resource.
Information about a resource from which the described resource is derived.
Name of the responsible corporation or organisation.
Country of the postal address.
Address of the electronic mailbox of the responsible party.
Telephone number(s) of a facsimile machine for the responsible party.
City, suburb, town or other locality of the postal address.
Name of the responsible person.
Physical location of the responsible party.
Position of the responsible person.
Address line for the postal address.
Postal code of the postal address.
Function performed by the responsible party in relation to the described resource.
Sector of the agent.
State or Territory of the postal address.
Telephone number(s) at which the responsible party may be contacted.
Location (address) for online access using a Uniform Resource Indicator.
Name of the responsible corporation or organisation.
Cost of obtaining the resource.
Country of the postal address.
Address of the electronic mailbox of the responsible party.
Telephone number(s) of a facsimile machine for the responsible party.
Time period (including time zone) when individuals can contact the responsible party.
Supplemental instructions about accessing the resource.
City, suburb, town or other locality of the postal address.
Name of the responsible person.
Information about physical access to premises and services for people with disabilities.
Physical address line of the responsible party.

Position of the responsible person.
Postal address of the responsible party.
Postal code of the physical address.
Function performed by the responsible party in relation to the described resource.
Sector of the agent.
State or Territory of the physical address.
Telephone number(s) at which the responsible party may be contacted.
Location (address) for online access using a Uniform Resource Indicator.
The date (and optionally time) that the metadata record was created. It is not the date the
Unique identifier for the metadata record.
The metadata character set is the code for the character set used in the metadata record. This pr
information. This property does not convey details about the individual, organisation and/or position associated with the resource itself.
The written language used for completing the metadata record.This property does not describe
The metadata standard followed for creation of the metadata.
The version of the metadata standard followed for creation of the metadata.
The date (and optionally time) that the metadata was last updated or modified. It is not the

Obligation
Mandatory
Mandatory unless a related property is used
Optional
Optional—may be used in place of date
Optional
Optional
Optional—may be used in place of date
Optional—may be used in place of date
Optional—may be used in place of date.
Mandatory
Optional
Conditional—Mandatory for descriptions of offline resources
Conditional—Mandatory for online resources
Optional
Conditional—Mandatory for information resources (optional for descriptions of services)
Recommended
Recommended if subject is not used
Recommended where the language of the resource is not English.
Recommended if function is not used.
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional

Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional

Optional
Optional
Optional
Optional
Optional
Optional
Optional
Optional
Mandatory
Mandatory
Conditional: It is not necessary to complete this property if the value is the default (utf-8) and/or the character encoding attribute is provided in an XML declaration.
Optional
Mandatory unless the value is the default.
Mandatory unless the value is the default
Mandatory
Optional

0.1 Legal environment
0.2 Resources
0.3 Relevance
0.4 Quality management
1.1 Professionalism
1.2 Transparency
1.3 Ethical standards
2.1 Concepts and definitions
2.2 Scope
2.3 Classification/sectorization
2.4 Basis for recording
3.1 Source data
3.2 Assessment of source data
3.3 Statistical techniques
3.4 Data validation
3.5 Revision studies
4.1 Periodicity and timeliness
4.2 Consistency

4.3 Revision
5.1 Data
5.2 Metadata
5.3 Assistance to users

https://dsbb.imf.org/sddsplus/dq

0. Prerequisites
0.1.1 Responsibility for collecting, processing, and disseminating statistics
0.1.2 Data sharing and coordination among data producing agencies
0.1.3 Confidentiality of individual reporters' data
0.1.4 Ensuring statistical reporting
0.2.1 Staff, facilities, computing resources, and financing
0.2.2 Ensuring efficient use of resources
0.3.1 Monitoring user requirements
0.4.1 Quality policy
0.4.2 Quality monitoring
0.4.3 Quality planning
1. Integrity
1.1.1 Impartiality of statistics
1.1.2 Selection of sources, methodology, and modes of dissemination
1.1.3 Commenting on erroneous interpretation and misuse of statistics
1.2.1 Disclosure of terms and conditions for statistical collection, processing, and dissemination
1.2.2 Internal governmental access to statistics prior to release
1.2.3 Attribution of statistical products
1.2.4 Advance notice of major changes in methodology, source data, and statistical techniques
1.3.1 Guidelines for staff behavior
2. Methodology
2.1.1 Concepts and definitions
2.2.1 Scope
2.3.1 Classification/sectorization
2.4.1 Valuation
2.4.2 Recording basis
2.4.3 Grossing/netting procedures
3. Accuracy and reliability
3.1.1 Source data collection programs
3.1.2 Source data definitions, scope, classifications, valuation, and time of recording
3.1.3 Source data timeliness
3.2.1 Source data assessment
3.3.1 Source data statistical techniques
3.3.2 Other statistical procedures
3.4.1 Validation of intermediate results
3.4.2 Assessment of intermediate data
3.4.3 Assessment of discrepancies and other problems in statistical outputs
3.5.1 Revision studies and analyses
4. Serviceability
4.1.1 Periodicity
4.1.2 Timeliness
4.2.1 Internal consistency
4.2.2 Temporal consistency
4.2.3 Intersectoral and cross-domain consistency

4.3.1 Revision schedule
4.3.2 Identification of preliminary and/or revised data
4.3.3 Dissemination of revision studies and analyses
5. Accessibility
5.1.1 Statistical presentation
5.1.2 Dissemination media and format
5.1.3 Advance release calendar
5.1.4 Simultaneous release
5.1.5 Dissemination on request
5.2.1 Dissemination of documentation on concepts, scope, classifications, basis of recording,
5.2.2 Disseminated level of detail
5.3.1 Dissemination of information on contact points
5.3.2 Availability of documents and services catalogs

af-base/country/SWE/category/IND00

Label
Abstract
Access Rights
Accrual Method
Accrual Periodicity
Accrual Policy
Alternative Title
audience
Date Available
Bibliographic Citation
Conforms To
contributor
coverage
Date Created
creator
date
Date Accepted
Date Copyrighted
Date Submitted
description
Audience Education Level
Extent
format
Has Format
Has Part
Has Version
identifier
Instructional Method
Is Format Of
Is Part Of
Is Referenced By
Is Replaced By
Is Required By
Date Issued
Is Version Of
language
License
Mediator
Medium
Date Modified
Provenance
publisher

References
Relation
Replaces
Requires
Rights
Rights Holder
source
Spatial Coverage
subject
Table Of Contents
Temporal Coverage
title
type
Date Valid

Definition
A summary of the resource.
Information about who access the resource or an indication of its security status.
The method by which items are added to a collection.
The frequency with which items are added to a collection.
The policy governing the addition of items to a collection.
An alternative name for the resource.
A class of agents for whom the resource is intended or useful.
Date that the resource became or will become available.
A bibliographic reference for the resource.
An established standard to which the described resource conforms.
An entity responsible for making contributions to the resource.
The spatial or temporal topic of the resource, the spatial applicability of the resource, or the juris
Date of creation of the resource
An entity primarily responsible for making the resource.
A point or period of time associated with an event in the life of the resource
Date of acceptance of the resource.
Date of copyright of the resource.
Date of submission of the resource.
An account of the resource.
A class of agents, defined in terms of progression through an educational or training context, fo
The size or duration of the resource.
The file format, physical medium, or dimensions of the resource.
A related resource that is substantially the same as the pre-existing described resource, but in a
A related resource that is included either physically or logically in the described resource.
A related resource that is a version, edition or adaptation of the described resource.
An unambiguous reference to the resource within a given context.
A process, used to engender knowledge, attitudes and skills, that the described resource is desi
A pre-existing related resource that is substantially the same as the described resource, but in a
A related resource in which the described resource is physically or logically included.
A related resource that references, cites or otherwise points to the described resource.
A related resource that supplants, displaces or supersedes the described resource.
A related resource that requires the described resource to support its function, delivery or cohe
Date of formal issuance (e.g. publication) of the resource
A related resource of which the described resource is a version, edition or adaptation.
A language of the resource.
A legal document giving official permission to do something withthe resource.
An entity that mediates access to the resource.
The material or physical carrier of the resource.
Date on which the resource was changed.
A statement of any changes in ownership and custody of the resource since its creation that are
An entity responsible for making the resource available.

A related resource that is referenced, cited, or otherwise pointed to by the described resource
A related resource
A related resource that is supplanted, displaced, or superseded by the described resource.
A related resource that is required by the described resource to support its function, delivery, or
Information about rights held in and over the resource.
A person or organization owning or managing rights over the resource.
Information about a resource from which the described resource is derived.
Spatial characteristics of the resource.
The topic of the resource.
A list of subunits of the resource.
Temporal characteristics of the resource.
A name given to the resource.
The nature or genre of the resource.
Date (often a range) of validity of a resource.

sdiction under which the resource is relevant.

r which the described resource is intended

nother format.

gned to support
nother format.

ence.

significant for its authenticity, integrity, and interpretation coherence

Contact person and organisation
Data source(s) used
Name ofcollection /source used
Direct source
Source Periodicity
Source metadata
Date last input received from source
Unit of measureused
Power code
Variables collected
Sampling
Periodicity
Referenceperiod
Base period
Date lastupdated
Releasecalendar
Contact person
Statisticalpopulation
Geographiccoverage
Sector coverage
Institutionalcoverage
Item coverage
Populationcoverage
Productcoverage
Other coverage
Key statisticalconcepts used
Classification(s)used
Aggregation & consolidation
Estimation
Imputation
Transformations
Validation
Index type
Weights
Seasonaladjustment
Othermanipulation &adjustments
OECDDisseminationformat(s)
Relatedpublishing

Contact person, title, unit,organisation, phone number,fax, number, email, city,country, postal code [nonMCV]
List original data source(s)used (administrative data,household survey,enterprise/establishmentsurvey, etc).
Refers to full title of theoriginal survey collection,administrative source,database or publication fromwhere the data were obtained
Refers to the source fromwhere the data was directlycollected
The time distance betweenobservations in source(whether stock or flow).Values: Yearly, Quarterly,monthly, irregular
Reference or link to metadatafrom source
Refers to the date on whichthe data was last receivedfrom the source, e.g. nationalagency or internationalorganisation. [non-MCV
Refers to the unit in whichassociated values aremeasured, e.g. USD [nonMCV]
Power of 10 by which thereported statistics should bemultiplied, e.g. “6” indicatingmillions of USD. [non-MCV]
List of variables collected or provision of questionnaire [non-MCV]
Refers to information onsample size, sample frame,sample updating, sample(other)
The time distance betweenobservations (whether stockor flow). Values: Yearly,Quarterly, monthly, irregular<other?>
Period of time the data referto.For business tendency orconsumer opinion surveys thisfield could also refer to theforecasting horizo
The period of time for whichdata used as the base of anindex number, constant pricesdata or other ratio, have beencollected.
Refers to the date on whichthe data was last updated.[non-MCV]
Refers to a general statementon the schedule of release ofdata.
OECD contact person, title,unit, phone number, number,email [non-MCV]
Target population (thestatistical universe aboutwhich information is sought).
The geographic area coveredby the data. [non-MCV]
The range of sectors covered by the data [non-MCV]
The range of institutionscovered by the data [nonMCV]
The range of items covered bythe data [non-MCV]
The population covered by thedata [non-MCV]
The range of products coveredby the data [non-MCV]
Other issues and informationconcerning the coverage ofthe data [non-MCV]
A statistical concept is a statistical characteristic of a time series or an observation. This item should define key statistical concepts included in the domain of study
A classification is a set of discrete, exhaustive and mutually exclusive observations which can be assigned to one or more variables to be measured in the collation and/or presentation of data. This item should list the name of all classifications
Aggregation is the combination of related categories, usually within a common branch of a hierarchy, to provide information at a broader level to that at which detailed observations are taken
Estimation is concerned withinference about the numericalvalue of unknown populationvalues from incomplete datasuch as a samp
Refers to procedures forentering a value for a specificdata item where the responseis missing or unusable.
Mention of interpolations, andother transformations,indicating method usedincluding, if relevant, formulasemployed for transforma
A procedure which provides, by reference to independent sources, evidence that an enquiry is free from bias or otherwise conforms to its declared purpose. It may be applied to a sample investigation with the object of showing that the sample is reasonably representative of the population and that the information collected is accurate. Refers to processes applied for the
Index type
Refers to information on sources of weights, nature of weights, period of current index weights, frequency of weight updates, weig
Seasonal adjustment is a statistical technique to remove the effects of seasonal calendar influences operating on a series. Seasonal effects usually reflect the influence of the seasons themselves either directly or through production series related to them, or social conventions. Should provide information to enable users to make an assessment of the validity of the seasonal adjustment applied. Such information would comprise: a short description of the method (software) used; the main parameters of the adjustment (e.g. additive v. multiplicative decomposition) and some of the derived information (e.g. trading-day
Manipulation and adjustments not mentioned under the headings Aggregation & consolidation, Estimation, Imputation, Validation, Index type, Weights, Sampling, Seasonal adjustment
Refers to the different dissemination media used to disseminate the data, e.g. news release, paper publication, on-line or database, CD-ROM or other. [non-MCV]
Gives links or references to web sites and publications where the data has been published, used for analytical purposes, etc

.[non-MCV]

]

on. [nonMCV].

ple.

ation

hts (other)

√

Government of India

Ministry of Statistics and Programme Implementation

Data Dissemination: National Metadata Structure (NMDS) For Statistical Products

August 2021

Government of India
Ministry of Statistics and Programme Implementation
Policy Implementation & Monitoring Division (PIMD)
Sardar Patel Bhawan, 4th Floor
Sansad Marg, New Delhi – 110001
Ph: (011) 23341867

Data Dissemination: National Metadata Structure (NMDS) For Statistical Products

August 2021

Introduction

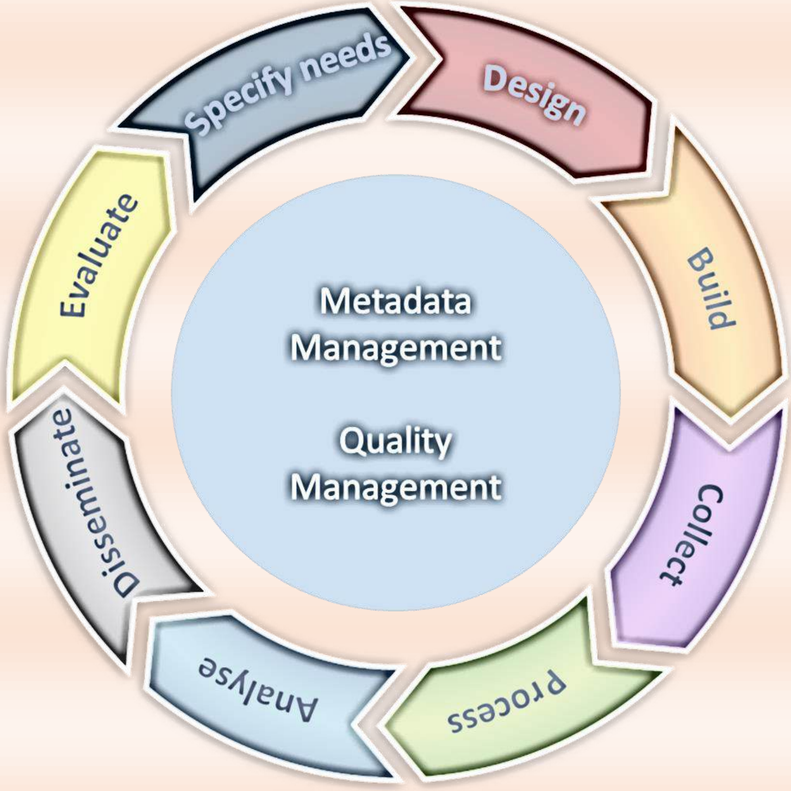
National Statistical Office (NSO), Ministry of Statistics & Programme Implementation, presents and disseminates data and metadata through different products like Census data (Economic Census); Survey data such as NSS Surveys viz. Household Surveys, etc, Annual Survey of Industries (ASI), Consumer Price Indices (CPI), and macro-economic aggregates like National Income, Index of Industrial Production (IIP). In addition, statistical data is presented in analytical publications such as NSS Reports, Annual Survey of Industries Reports, National Indicator Framework (NIF) for monitoring SDGs, Energy Statistics, EnviStats India, Women & Men in India etc., which provide analysis of data, supported by the visual presentation of that data in the form of graphs and maps.

The production of data and presentation of metadata structure requires an overview of the arrangements, technical infrastructure and skills required for a holistic and integrated approach to the presentation and dissemination of statistical data and metadata to different user groups. National Metadata Structure (NMDS) is to provide guidelines for the data producer to adhere to a basic minimum quality standard in order to establish and maintain the quality of data and enhance ease in sharing data. The specific objectives of this document are:

- to promote reporting for each type of statistical process and its outputs across different Ministries/Divisions/Departments of NSO, hence facilitating comparisons across processes and outputs;
- to ensure that producer reports contain all the information required to facilitate identification of quality issues and potential improvements in statistical processes and their outputs; and
- to ensure that user reports contain all the information required by users to assess whether statistical outputs are fit for the purposes they have in mind.

A. What is Metadata?

A.1. Metadata should contain all the information users need to analyse a dataset and draw conclusions. It increases data accessibility by summarizing the most important information (i.e. methodology, sampling design, interview mode, etc.) required for analyzing a dataset which alleviates the need for users to search for supporting documents and reports. Furthermore, good metadata clearly articulates the potential uses for a dataset, preventing potential misuses. Metadata is also a tool for rendering complex microdata structures into something meaningful, navigable, and user-friendly. Finally, the adoption of well-known metadata schemas and vocabularies allows for semantic interoperability.


Specify needs
Design
Build
Collect
Process
Analyse
Disseminate
Evaluate
Metadata Management
Quality Management

The Metadata process is fully integrated in the Generic Statistical Business Process Model[1] (GSBPM) which has metadata as one of the key elements in the version 5.1.

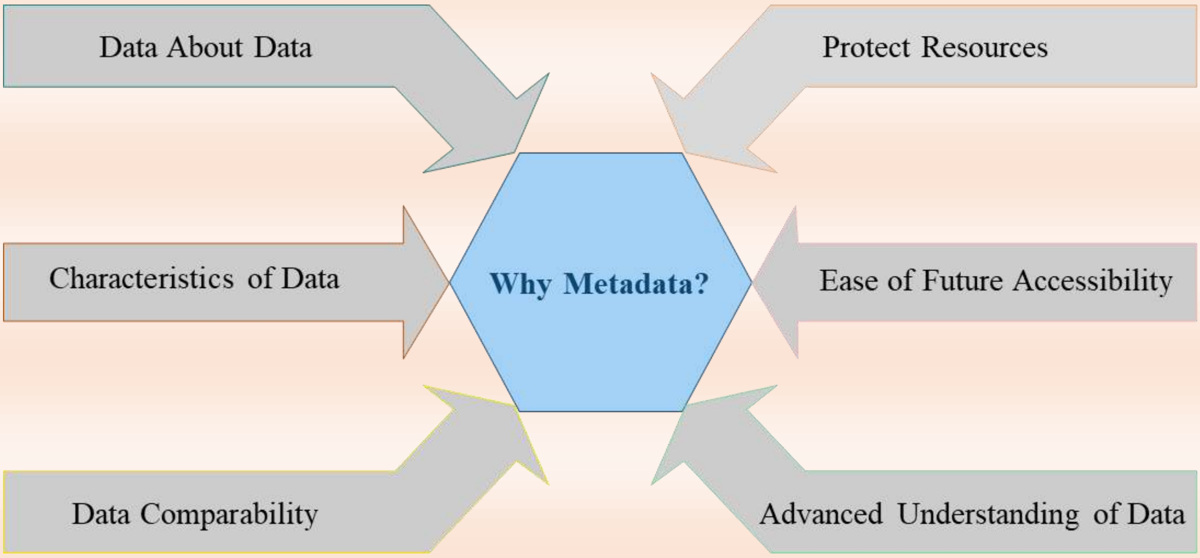
---

[1] UNECE: United Nations Economic Commission for Europe, https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1

B. Why Metadata?

B.1. In most information technology usages, the prefix of meta conveys “an underlying definition or description.” So it is that, at its most basic, metadata is data about data. More precisely, however, metadata describes data containing specific information like type, length, textual description and other characteristics. Metadata makes it much easier to find relevant data and to use a dataset, users need to understand how the data is structured, definitions of terms used, how it was collected, and how it should be read.

B.2. Metadata is an important way to protect resources and their future accessibility. For archiving and preservation purposes, it takes metadata elements that track the object’s lineage, and describe its physical characteristics and behaviour so it can be replicated on technologies in the future.


Data About Data
Protect Resources
Characteristics of Data
Why Metadata?
Ease of Future Accessibility
Data Comparability
Advanced Understanding of Data

B.3. In today’s modern data driven world and in the era of digital transactions, huge amount of data is generated on real time basis, and lately, a large number of organisations/agencies have started producing data, the quantum of which is huge, and thus arises a need of standard regulatory framework to be laid down to assure the quality of data produced by different producers. It will also serve the purpose of ensuring data comparability across time horizons so as to enable better understanding of different social and economic movements.

B.4. Although metadata may not seem exciting or impressive, the true importance of metadata can never be underrated and hence, is important to take a concerted effort to build sound metadata structure to draw maximum gains from varied data sets.

C. Role of MoSPI in Building Metadata Structure

C.1. MoSPI being a nodal agency for planned development of the statistical system in the country is also responsible for maintaining the highest standards of data quality which adhere to basic guidelines of International Agencies so as to ensure India’s statistical system is one of the frontrunners in quality data producer. MoSPI aims at raising the National Statistical System (NSS) to the epitome of being one of the best professionally equipped government data producing agencies by building the best of IT infrastructure among others in the system, and Metadata is one of the building blocks to achieve the objective.

C.2. The document presents the NMDS in two formats- the first one is the indexed version as NMDS concepts (Section F), and the second one presents details of concepts through definition and guidelines (Section G).

D. Metadata Management

D.1 It is advisable to put in place a metadata policy by the official statistical producing agencies, ab initio. The policy is a set of broad, high level principles that form the guiding framework within which metadata management can operate.

D.2 Once the metadata policy is put in place, for an organisation, metadata should be compiled and maintained actively. Otherwise, the currency, and thus use of Metadata will degrade with time. To realise the full capabilities of Metadata, it is necessary that the Metadata are maintained over a long period of time. Even with investment in technically sophisticated search tools, such systems may find little stakeholders acceptance, if the data are incomplete or is not updated regularly.

While preparing the NMDS, the following core principles should be borne in mind:

i. Metadata Handling:
    a. Statistical Business Process Model
    b. Active, not Passive
    c. Reuse for Efficiency
    d. Version Preservations

ii. Metadata Authority
    e. Registration
    f. Single Source
    g. One Entry/Update
    h. Standards Variations

iii. Relationship to Statistical Business Processes
    i. Integrity
    j. Matching Metadata
    k. Describe Flow
    l. Capture at Source
    m. Exchange and Use

iv. Users
    n. Identify Users
    o. Variant Formats
    p. Availability

E. Retention, Preservation, and Destruction

National Statistics constitute valuable and irreplaceable assets whose value can increase through widespread and long-term use. National Statistics should thus be backed by the Data Management Policy eliciting the arrangements it has in place for the retention, long term preservation, and destruction of its resources including Metadata.

F. National Metadata Structure (NMDS) Concepts - Index

Item No	Concept name
1	**Contact**
1.1	Contact Organisation
1.2	Compiling Agency
1.3	Contact Details
2	**Statistical Presentation and Description**
2.1	Data description
2.2	Classification system
2.3	Sector coverage
2.4	Statistical concepts and definitions
2.5	Statistical unit
2.6	Statistical population
2.7	Reference Period
2.8	Data Confidentiality
3	**Institutional Mandate**
3.1	Legal acts and other agreements
3.2	Data sharing
3.3	Release policy
3.4	Release calendar
3.5	Frequency of dissemination
3.6	Data access
4	**Quality Management**
4.1	Documentation on methodology
4.2	Quality documentation
4.3	Quality assurance
4.4	Quality assessment
5	**Accuracy and Reliability**
5.1	Sampling error
6	**Timeliness**
6.1	Timeliness
7	**Coherence and Comparability**
7.1	Comparability – over time
7.2	Coherence
8	**Statistical Processing**
8.1	Source data type
8.2	Frequency of data collection
8.3	Data collection method
8.4	Data validation
8.5	Data compilation
9	**Metadata Update**
9.1	Metadata last posted
9.2	Metadata last update

G. Details of NMDS Concepts

Item No	Concept name	Definition	Guidelines
1	**Contact**	**Individual or organisational contact points for the data or metadata, including information on how to reach the contact points.**	
1.1	Contact Organisation	The name of the organisation of the contact points for the data or Metadata.	Provide the full name (not just acronym/code name) of the organisation responsible for the processes and outputs (data and metadata) that are the subject of the report
1.2	Compiling agency	Organisation collecting and/or elaborating the data being reported	Provide the full name of the Department/Division under the organisation responsible for the processes and outputs (data and metadata) that are the subject of the report
1.3	Contact Details	The details of the contact points for the data or metadata.	Provide contact details of contact point(s) in following format: a. Name of Organisation owning the processes and outputs b. Author (if different from (a)) c. Disseminating Agency (if different from (a) and (b)) d. Name (first and last names) e. Designation f. Postal address g. email address (preferably designation based) h. Contact number i. Fax number If more than one name is provided, the details of main contact should be indicated. If the author of the report is different from the person(s) responsible for process and its outputs, provide this name also with his/her details

Item No	Concept name	Definition	Guidelines
2	**Statistical Presentation and Description**	**Description of the disseminated data which can be displayed to users as tables, graphs or maps**	
2.1	Data description	Main characteristics of the data set, referring to the data and indicators disseminated.	Describe briefly the main characteristics of the data in an easily and quickly understandable manner, referring to the main variables disseminated.
2.2	Classification system	Arrangement or division of objects into groups based on characteristics which the objects have in common	List all classifications and breakdowns that are used in the data (with their detailed names) and provide links (if publicly available). Type of dis-aggregation available in the data sets - for example rural-urban, male-female, etc. and whether data is available at the sub-national level or not, should be clearly specified.
2.3	Sector coverage	Main economic or other sectors	List the main economic or other sectors covered by the data and the size classes used, for example, Health/ Education/ Manufacturing etc for sectors and classes based on number of employees for size classes
2.4	Statistical concepts and definitions	Statistical characteristics of statistical observations, variables	Define and describe briefly the main statistical variables that have been observed or derived. Indicate their types.
2.5	Statistical unit	Entity for which information is sought and for which statistics are ultimately compiled.	Define the type of statistical unit about which data are collected, e.g. enterprise, household, etc.
2.6	Statistical population	The total population of a defined class of people, objects or events	Define the target population of statistical units for which information is sought. For example, agricultural household, general household, industrial unit, etc.

Item No	Concept name	Definition	Guidelines
2.7	Reference Period	The length of time for which data are available	State the time period(s) for which data is collected
2.8	Data Confidentiality[2]	Rules applied for treating the datasets to ensure statistical confidentiality and prevent unauthorised disclosure.	Describe the procedures that are used in protecting confidentiality, viz., anonymisation, legal provision, if any.
3	**Institutional Mandate**	**Law, set of rules or other formal set of instructions assigning responsibility as well as the authority to an organisation for the collection, processing, and dissemination of statistics**	
3.1	Legal acts and other agreements	Legal acts or other formal or informal agreements that assign responsibility as well as the authority to an agency for the collection, processing, and dissemination of statistics	State the national legal acts and/or other reporting agreements
3.2	Data sharing	Arrangements or procedures for data sharing and coordination between data producing agencies.	Describe the arrangements, procedures or agreements to facilitate data sharing and exchange between data producing agencies within the national statistical system

---

[2] All statistical information published by any agency shall be arranged in such a manner so as to prevent any particulars becoming identifiable by any person (other than the informant by whom those particulars were supplied) as the particulars relating to the informant who supplied it, even through the process of elimination (Source: Collection of Statistics Act, 2008).

Item No	Concept name	Definition	Guidelines
3.3	Release policy	Rules for disseminating statistical data to all interested parties	State if the release of the products is governed by some policy etc.
3.4	Release calendar	The schedule of statistical release dates.	State whether there is a release calendar for the statistical outputs from the process being reported, and if so, whether this calendar is publicly accessible and if yes, give a link or reference.
3.5	Frequency of dissemination	The time interval at which the statistics are disseminated over a given time period.	State the frequency with which the data are disseminated, e.g. monthly, quarterly, yearly.
3.6	Data access	The conditions and modalities by which users can access, use and interpret data	State the conditions and link on website from where the user can access the data For easy access of users, following details should also be mentioned about the dataset: **Title:** Name by which the data is known **Dataset Edition:** Edition of data (ex: first, second, final etc) **Dataset Reference data type:** Type of data entered in the field (ex: .txt, .dbf, .xls) **Presentation Format:** Presentation format of the data (ex: document, map, table, etc.) **Dataset Language:** language of any text in the data **Status/Version:** How updated is the data?

Item No	Concept name	Definition	Guidelines
4	**Quality Management**	**Systems and frameworks in place within an organisation to manage the quality of statistical products and processes.**	
4.1	Documentation on methodology	Descriptive text and references to methodological documents available.	List reference metadata files, methodological papers, summary documents and handbooks relevant to the statistical process.
4.2	Quality documentation	Documentation on procedures applied for quality management and quality assessment.	List relevant quality related documents, for example, other quality reports, studies.
4.3	Quality assurance	All systematic activities implemented that can be demonstrated to provide confidence that the processes will fulfil the requirements for the statistical output.	Describe the procedures (such as use of a general quality management system based on ISO 9000 series) to promote general quality management principles in the organisation. Describe the quality assurance framework used to implement statistical quality principles. Describe the quality assurance procedures specifically applied to the statistical process for which the report is being prepared, for example training courses, process monitoring, benchmarking, assessments, and use of best practices. Describe any ongoing or planned improvements in quality assurance procedures.

Item No	Concept name	Definition	Guidelines
4.4	Quality assessment	Overall assessment of data quality, based on standard quality criteria.	Summarise the results of the most recent quality assessments and cross reference to the chapters in the report where the results are presented in more detail.
5	**Accuracy and Reliability**	**Accuracy of data is the closeness of computations or estimates to the exact or true values that the statistics were intended to measure. Reliability of the data, defined as the closeness of the initial estimated value to the subsequent estimated value.**	
5.1	Sampling error	That part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a subset of the population is enumerated.	If probability sampling is used: • for user reports-provide the range of variation of the A1[3] indicator among key variables at user report level of detail; • for producer reports-provide the range of variation of the A1 indicator among key variables at producer report level of detail; • indicate the impact of sampling error on the overall accuracy of the results; • state how the calculation of sampling error is affected by adjustments for nonresponse, misclassifications and other sources of uncertainty, such as outlier treatment. If non-probability sampling is used: provide an assessment of representativeness, a motivation for the invoked model for estimation and risk of sampling bias

---

[3] In international terminology, A1 is simply used for Sampling Errors. In other words, indicator for Sampling Errors is denoted by A1. For further details, refer to: https://ec.europa.eu/eurostat/documents/64157/4373903/02-ESS-Quality-and-performance-Indicators-2014.pdf/5c996003-b770-4a7c-9c2f-bf733e6b1f31

Item No	Concept name	Definition	Guidelines
6	**Timeliness**	**The timeliness of the data collection release to be compiled.**	
6.1	Timeliness	Length of time between data availability, the event or phenomenon the data describe, and final release to its users.	Outline the reasons for the time lag, if any. Outline efforts to reduce time lag in future.
7	**Coherence and Comparability**	**Adequacy of statistics to be reliably combined in different ways and for various uses and the extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics**	
7.1	Comparability – over time	The extent to which statistics are comparable or reconcilable over time	Provide information on possible limitations in the use of data for comparisons over time. Distinguish three broad possibilities: 1. There have been no changes, in which case this should be reported. 2. There have been some changes but not enough to warrant the designation of a break in series. 3. There have been sufficient changes to warrant the designation of a break in series.
7.2	Coherence	The extent to which statistics are reconcilable with System of National Accounts	For producer reports only. Where relevant, the results of comparisons with the System of National Account framework / Other Statistical Standards and feedback from System of National Accounts / Other Statistical Standards with respect to coherence and accuracy problems should be reported and should be a trigger for further investigation.

Item No	Concept name	Definition	Guidelines
8	**Statistical Processing**	**Any statistical processing undertaken to finalise the data**	
8.1	Source data type	Characteristics and components of the raw statistical data used for compiling statistical aggregates.	Indicate if the data are based on a survey, administrative data, multiple data sources, or macro-aggregates. In the event of multisource or macro-aggregate processes describe each source. For each dataset from an administrative source, summarise the source, its primary purpose, and the most important data items acquired
8.2	Frequency of data collection	Time interval at which the source data are collected	Indicate the frequency of data collection (e.g. monthly, quarterly, annually, or continuous).
8.3	Data collection method	Method applied for gathering data for official statistics	For each source of survey data: • describe the method(s) used to gather data from respondents; • annex or hyperlink the questionnaires(s). For each source of administrative data: • describe the acquisition process and how it was tested. For all sources: • describe the types of checks applied at the time of data entry.

Item No	Concept name	Definition	Guidelines
8.4	Data validation	Process of monitoring the results of data compilation and ensuring the quality of statistical results	Describe the procedures for checking and validating the source data and how the results are monitored and used. Describe the procedures for validating the aggregate output data (statistics) after compilation, including checking coverage and response rates, and comparing with data for previous cycles and with expectations. List other output datasets to which the data relate and outline the procedures for identifying inconsistencies between the output data and these other datasets
8.5	Data compilation	Operations performed on data to derive new information according to a given set of rules.	Describe the procedures for imputation, the most common reasons for imputation and imputation rates within each of the main strata. Describe the likely impact of imputation. Describe the procedures to derive new variables and to calculate aggregates and complex statistics. Describe the procedures for adjustment for non-response and the corrections to the design weights to account for differences in response rates. Describe the calculation of design weights, including calibration (if used). Describe the procedures for combining input data from different sources.

Item No	Concept name	Definition	Guidelines
			Provide the ratio of the number of replaced values to the total number of values for a given variable. Specific reference to formula shall be made. The formula or mathematical equation used while computing different variables in the report may be described here in a structured format showing the Numerator; Denominator and Multiplier used for computing the same
9	**Metadata Update**	**The date on which the metadata element was inserted or modified in the database.**	
9.1	Metadata last posted	Date of the latest dissemination of the metadata	The date when the complete set of metadata was last disseminated as a block should be provided (manually, or automatically by the metadata system).
9.2	Metadata last update	Date of last update of the content of the metadata.	The date when any metadata were last updated should be provided (manually, or automatically by the metadata system).

ಕರ್ನಾಟಕ ಸರ್ಕಾರ
Government of Karnataka

Conference of Central and State Statistical Organizations(CoCSSO)
New Delhi

State Participation in conducting National Sample Survey

Directorate of Economics & Statistics,Karnataka

Date:12-08-2024

State’s Participation

- Till the 8th round (1954-55), only GOI was involved in survey work.
- Karnataka has been participating in the National Sample Surveys since 1963 i.e. 18th round on an equal matching basis.
- The main idea of State participation in the NSS Programme is to provide more desegregated results at the regional level by pooling the results related to central and state samples.

FIELD WORK

- **Sample design, selection of sample, preparation of questionnaire, instructions, etc., are the responsibility of the NSSO in respect State Sample**
- **NSSO also provides blank Schedules, training, technical guidance for conducting the survey.**
- **The State shall adopt uniform concepts, definitions and field procedures.**

FIELD WORK

- DES conduct the state level training for district staff for field work, data entry and first phase validation

- Two Statistical Inspectors (with minimal variation in some districts) and Assistant Statistical Officers in District Statistical Offices are trained for the purpose of carrying out field work.

DATA ENTRY & VALIDATION

- Data entry, Validation and Tabulation software is provided by the NSSO and is used in the State.
- In Karnataka, the responsibilities range from field operations to publication of survey results conducted by the Directorate of Economics and Statistics
- DES conduct the State level training for District Staff for data entry and first stage validation.

Processing of data & publication of reports

- DES brought out the publications related to state sample data.
- Since the 66th round, DES has been pooling the data of the central and state samples and brought out the publications of the pooled results.
- DES has published the pooled data results till the 76th round.
- State sample reports are hosted on the Department's website.

Pooling of Central and State Samples Data of NSS

- Two data sets (one from NSSO and the other by DES) were pooled together to obtain reliable and acceptable (free from agency bias) estimates.
- Pooling results in greater accuracy of estimates.
- To fill data gaps at district level on socio-economic indicators.
- Derivation of weighting diagram based on pooled data for construction of proposed CPI Numbers for rural and urban areas at sub-state level

National Sample Survey

- The 76th round has been completed and results have been published.
- For rounds 77 and 78, fieldwork and validation completed, pooling software is awaited.
- The 77th and 78th round results are yet to be declared.

National Sample Survey -79th Round

- **The present NSS 79th round of National Sample Survey is devoted for collection of data on Comprehensive Annual Modular Survey & AYUSH.**
- **Period of survey : July 2022 - June 2023**
- **Till 78th round State was conducting survey through Paper Assisted Personnel Interview (PAPI).**
- **79th Round: DES was conducted through Computer Assisted Personnel Interview (CAPI) mode.**
- **DES used the software provided by the NSSO.**

National Sample Survey -79th Round

- BY using CAPI in the field the survey is completed on time and it reduce time.
- Due to the inbuilt software the error is reduced and some data is validated automatically.
- DES has taken the steps to use CAPI for the 79th NSS Survey, which results in all samples i.e. 632 completed on time.
- Every data saved in CAPI, awaiting the software for validation, tabulation and pooling from NSSO

Last 8 Rounds subjects details

Round	Subject & Schedule	Samples allotted	Survey Period	Published Date/Month
70	Household Listing Schedule Rural & Urban Sch 0.0	336	2013 January - 2013 December	Pooling software was not provided by NSSO
	Land & Livestock Holdings Sch 33.1 (Rural Only)			
	Debt & Investments Sch 18.2			
	Situation Assement of Agricultural holdings Sch 33.1 (Rural only)			
71	Household Listing Schedule Rural & Urban Sch 0.0	370	2014 January - 2014 June	2 reports has been published on 25.07.2019 & 31.08.2019
	Social Consumption : Health Sch. 25.0			
	Social Consumption : Education Sch. 25.2			
72	Household Listing Schedule Rural & Urban Sch 0.0	596	2014 July - 2015 June	Report has been published
	Domestic Tourism Sch 21.1			
	Household Expenditure on Services & Durable Goods Sch 1.5			
	Household Consumer Expenditure Sch 1.60			

Last 8 Rounds subjects details

Round	Subject & Schedule	Samples allotted	Survey Period	Published Date/Month
72	Household Expenditure Consumer on Food Items Sch 1.61			
	Household Expenditure Consumer on Non Food Items Sch 1.62			
73	Household & Enterprises Listing Schedule Rural & Urban Sch 0.0	664	2015 July - 2016 June	Report has been published
	Unincorporated Non agricultural Enterprises (Excluding Construction) Sch 2.34			
74	Service Sector Enterprises (Economic Survey Frame) (1716 - BBMP)	2714	2016 July - 2017 June	Survey was dropped
75	Household Listing Schedule Rural & Urban Sch 0.0	596	2017 July - 2018 June	Report on Household Consumer Expenditure has not been published because of non availability of unit level data from NSSO & other two reports has been published
	Household Consumer Expenditure Sch 1.60			
	Household Social Consumption on Health Sch.25.0			
	Household Social Consumption on Education Sch 25.2			

Last 8 Rounds subjects details

Round	Subject & Schedule	Samples allotted	Survey Period	Published Date/Month
76	Household Listing Schedule Rural & Urban Sch 0.0	410	July 2018 – Dec 2018	Two Reports has been published
	Drinking Water, Sanitation, Hygiene and Housing condition Sch 1.2			
	Disability People in Karnataka Sch 26.0			
77	Household Listing Schedule Rural & Urban Sch 0.0	419	Jan 2019 – Dec 2019	Waiting for pooling software from NSSO
	Situation Assement of Agricultural holdings & Land & Live stock holdings of Households in Karnataka Sch 33.1 (Rural only)			
	Debt & Investments Sch 18.2			
78	Household Listing Schedule Rural & Urban Sch 0.0	716	Jan 2020 – Dec 2020	State level tabulation work is under progress.
	Expenditure on Domestic Tourism Sch 21.1 (Due to Covid pendmic survey has dropped)			
	Multiple indicators Sch 5.1			
79	Household Listing Schedule Rural & Urban Sch 00CM	632	July 2022 – june 2023	Field work completed
	Comprehensive Annual Modular Survey CAMS Sch CAMS 2022-23			
	Survey On AYUSH (Ayurveda, Yoga& Naturopathy , Unani, Siddha & Hpmeopathy) Sch AYUSH 2022-23			

Thank you

National Metadata Structure (NMDS)

A Standard Set of Guidelines for Ensuring Quality of Statistical Products Released

Objectives

- Promote reporting for each type of statistical process and its outputs across different Ministries/Divisions/Departments of NSO, hence facilitating comparisons across processes and outputs;

- Ensure that producer reports contain all the information required to facilitate identification of quality issues and potential improvements in statistical processes and their outputs; and

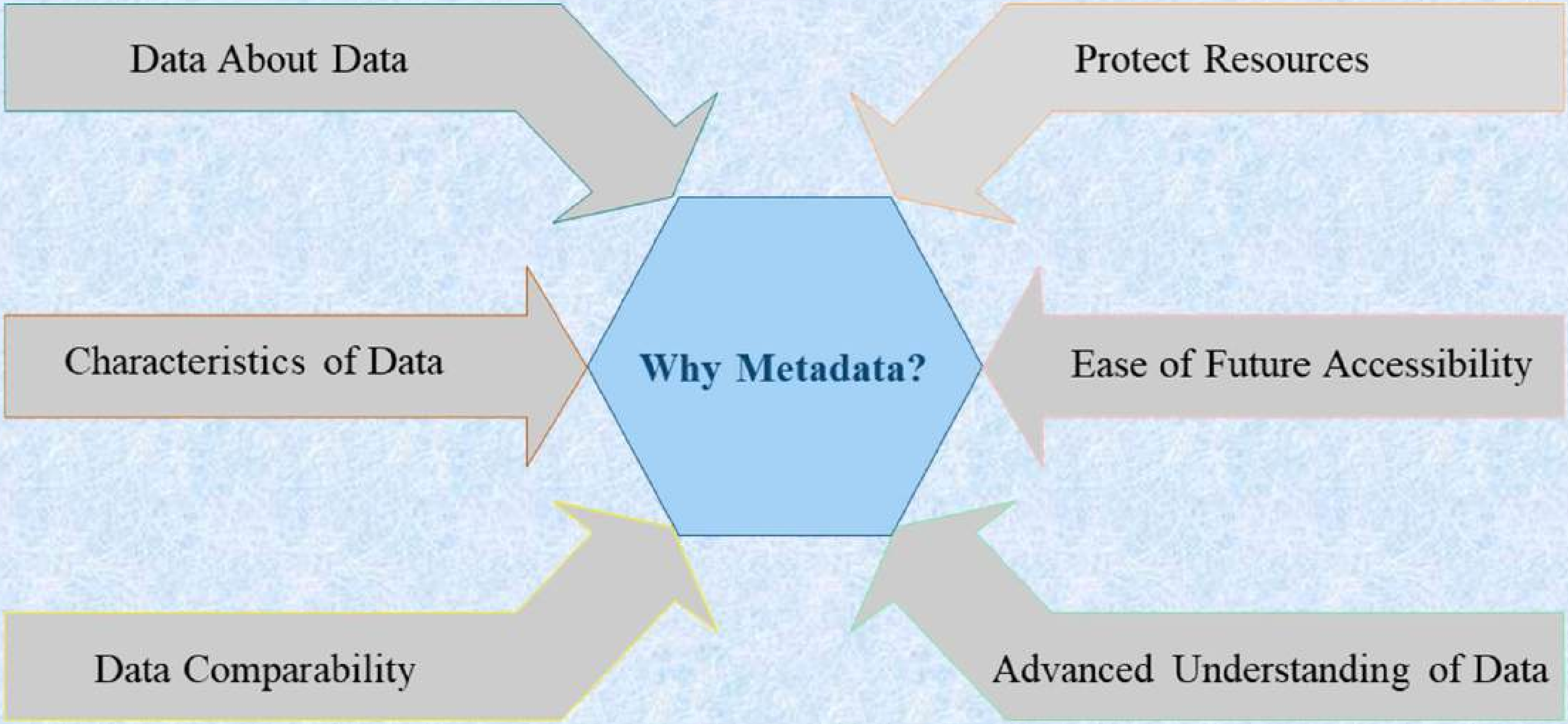
- Ensure that user reports contain all the information required by users to assess whether statistical outputs are fit for the purposes they have in mind.

What is metadata?

- It is structured information or documentation about data which:
  - informs users about the content, quality and condition of data;
  - describes the structure of datasets, explains, locates, or makes it easier to retrieve, use, or manage data;
  - provides information on the processes of data production;
  - guides on proper usage or interpretation of data.

- "Data that define and describe other data" (ISO definition)
- "Information about information" (Dion, 2006)

Why Metadata



Data About Data

Protect Resources

Characteristics of Data

Why Metadata?

Ease of Future Accessibility

Data Comparability

Advanced Understanding of Data

What does metadata do?

- Assist in retrieving and processing data
- Support correct use of data
- Provide transparency in data
- Enhance interoperability
- Improve archiving, preservation, institutional memory over time

Types of metadata

- **Structural metadata**
  - provide information about the structure of the dataset
  - act as identifiers and descriptors of the data, making it possible to properly identify, retrieve and browse the data
- **Reference metadata**
  - allow a thorough understanding and interpretation of the corresponding statistical data
  - Describe the concepts, definitions, methodology and quality of data; production and dissemination process, data access conditions, release policy, confidentiality, etc

Metadata standards (cont’d)

Two international metadata standards are becoming well established:

- SDMX (Statistical Data and Metadata Exchange)

  a number of international agencies have endorsed SDMX; supported by the UN Statistical Commission

-DDI (Data Dissemination Initiative)

  Microdata Management Toolkit of WB uses the DDI metadata standard

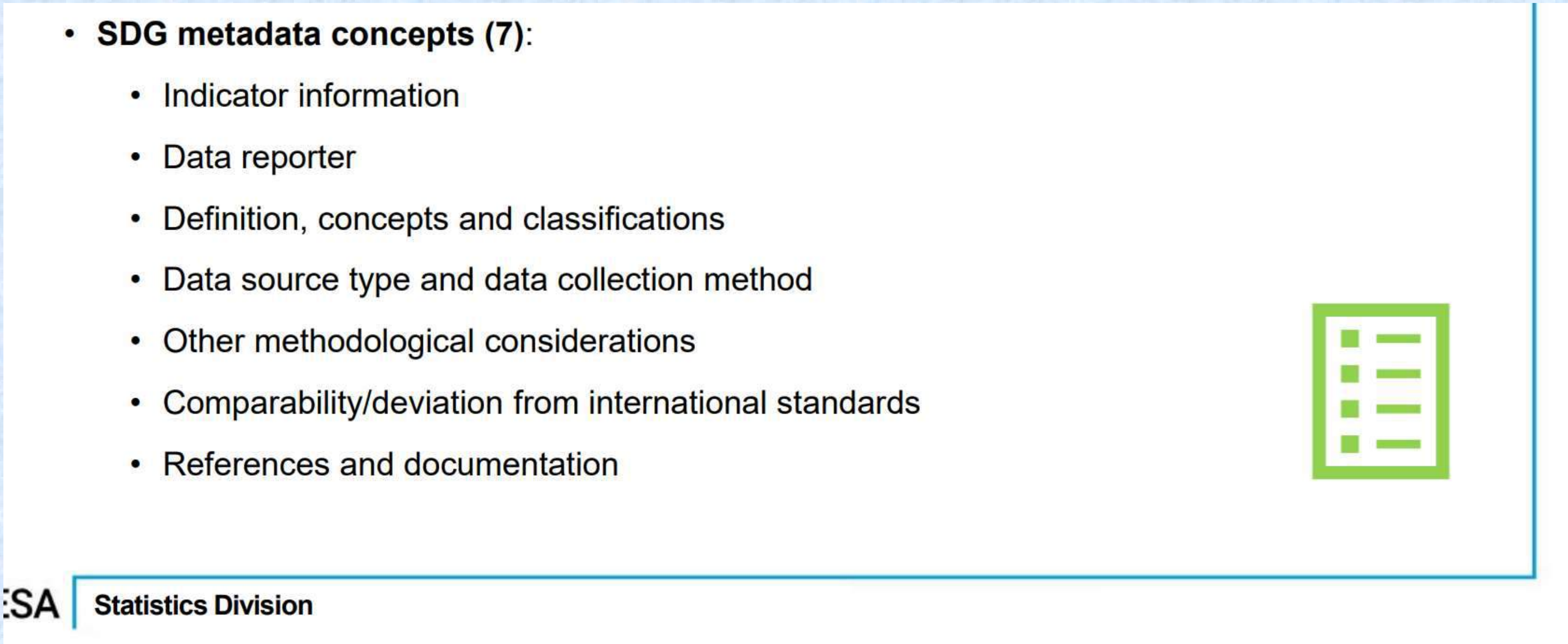
Users of metadata

Metadata support the knowledge of potential user of statistical information. The major users include:

- Users of statistical data
- Producers of statistical data
- Researchers on the development of statistical systems

Metadata Structure: International Organisations

UN SDG Metadata Concept



- **SDG metadata concepts (7)**:
    - Indicator information
    - Data reporter
    - Definition, concepts and classifications
    - Data source type and data collection method
    - Other methodological considerations
    - Comparability/deviation from international standards
    - References and documentation

ESA

Statistics Division

Figure 2. The Euro SDMX Metadata Structure

	Concept Name
1	**Contact**
1.1	Contact organisation
1.2	Contact organisation unit
1.3	Contact name
1.4	Contact person function
1.5	Contact mail address
1.6	Contact email address
1.7	Contact phone number
1.8	Contact fax number
2	**Metadata update**
2.1	Metadata last certified
2.2	Metadata last posted
2.3	Metadata last update
3	**Statistical presentation**
3.1	Data description
3.2	Classification system
3.3	Sector coverage
3.4	Statistical concepts and definitions
3.5	Statistical unit
3.6	Statistical population
3.7	Reference area
3.8	Time coverage
3.9	Base period
4	**Unit of measure**
5	**Reference period**
6	**Institutional mandate**
6.1	Legal acts and other agreements
6.2	Data sharing

	Concept Name
7	**Confidentiality**
7.1	Confidentiality - policy
7.2	Confidentiality - data treatment
8	**Release policy**
8.1	Release calendar
8.2	Release calendar access
8.3	User access
9	**Frequency of dissemination**
10	**Dissemination format**
10.1	News release
10.2	Publications
10.3	On-line database
10.4	Micro-data access
10.5	Other
11	**Accessibility of documentation**
11.1	Documentation on methodology
11.2	Quality documentation
12	**Quality management**
12.1	Quality assurance
12.2	Quality assessment
13	**Relevance**
13.1	User needs
13.2	User satisfaction
13.3	Completeness
14	**Accuracy and reliability**
14.1	Overall accuracy
14.2	Sampling error
14.3	Non-sampling error

	Concept Name
15	**Timeliness and punctuality**
15.1	Timeliness
15.2	Punctuality
16	**Comparability**
16.1	Comparability - geographical
16.2	Comparability - over time
17	**Coherence**
17.1	Coherence - cross domain
17.2	Coherence - internal
18	**Cost and burden**
19	**Data revision**
19.1	Data revision - policy
19.2	Data revision - practice
20	**Statistical processing**
20.1	Source data
20.2	Frequency of data collection
20.3	Data collection
20.4	Data validation
20.5	Data compilation
20.6	Adjustment
21	**Comment**

MoSPI’s Metadata Structure

Brief Concepts of Metadata Structure

Item No	Concept name	Item No	Concept name	Item No	Concept name
1	**Contact**	3.2	Data sharing	7.2	Coherence
1.1	Contact Organisation	3.3	Release policy	**8**	**Statistical processing**
1.2	Compiling Agency	3.4	Release calendar	9.1	Source data type
1.3	Contact Details	3.5	Frequency of dissemination	9.2	Frequency of data collection
2	**Statistical presentation and description**	3.6	Data access	9.3	Data collection method
2.1	Data description	**4**	**Quality management**	9.4	Data validation
2.2	Classification system	4.1	Documentation on methodology	9.5	Data compilation
2.3	Sector coverage	4.2	Quality documentation	**9**	**Metadata update**
2.4	Statistical concepts and definitions	4.3	Quality assurance	9.1	Metadata last posted
2.5	Statistical unit	4.4	Quality assessment	9.2	Metadata last update
2.6	Statistical population	**5**	**Accuracy and reliability**		
2.7	Reference Period	5.1	Sampling error		
2.8	Base period	**6**	**Timeliness**		
2.9	Data Confidentiality	6.1	Timeliness		
3	**Institutional mandate**	**7**	**Coherence and comparability**		
3.1	Legal acts and other agreements	7.1	Comparability – over time		

Elements of Metadata Structure: Comparison amongst selected NSOs

Detailed Concepts of Metadata

Contact

Back to Brief Concepts

Item No	Concept name	Definition	Guidelines
1	**Contact**	**Individual or organisational contact points for the data or metadata, including information on how to reach the contact points.**	
1.1	Contact Organisation	The name of the organisation for data or metadata.	Provide the full name (not just acronym) of the organisation responsible for the processes and outputs (data and metadata) that are the subject of the report
1.2	Compiling Agency	Organisation collecting and/or elaborating the data being reported	Provide the full name of the Division under the organisation responsible for the processes and outputs (data and metadata) that are the subject of the report
1.3	Contact Details	The details of the contact points for the data or metadata.	Provide contact details of contact point(s) in following format: a. Organisation Name owning the processes and outputs: b. Author (if different from (a)) c. Disseminating Agency (if different from (a) and (b)) d. Name (first and last names): e. Designation: f. Postal address: g. email address (preferably designation based): h. Contact number: i. Fax number: If more than one name is provided, the details of main contact should be indicated. If the author of the report is different from the person(s) responsible for process and its outputs, provide this name also with his/her details

Statistical Presentation and Description

Back to Brief Concepts

2	**Statistical presentation and description**	**Description of the disseminated data which can be displayed to users as tables, graphs or maps**	
2.1	Data description	Main characteristics of the data set, referring to the data and indicators disseminated.	Describe briefly the main characteristics of the data in an easily and quickly understandable manner, referring to the main variables disseminated.
2.2	Classification system	Arrangement or division of objects into groups based on characteristics which the objects have in common	List all classifications and breakdowns that are used in the data (with their detailed names) and provide links (if publicly available). Type of dis-aggregation available in the data sets - for example rural-urban, male-female, etc. and whether data is available at the sub-national level or not, should be clearly specified.
2.3	Sector coverage	Main economic or other sectors	List the main economic or other sectors covered by the data and the size classes used, for example, Health/ Education/ Manufacturing etc
2.4	Statistical concepts and definitions	Statistical characteristics of statistical observations, variables	Define and describe briefly the main statistical variables that have been observed or derived. Indicate their types.
2.5	Statistical unit	Entity for which information is sought and for which statistics are ultimately compiled.	Define the type of statistical unit about which data are collected, e.g. enterprise, household, etc.
2.6	Statistical population	The total population of a defined class of people, objects or events	Define the target population of statistical units for which information is sought
2.7	Reference Period	The length of time for which data are available	State the time period(s) for which data is collected
2.8	Base period	The period of time used as the base of an index number, or to which a constant series refers.	Note that this concept applies only to certain types of outputs, such as indexes, for which a base period is defined and used. Eg, CPI, IIP, etc.
2.9	Data Confidentiality	Rules applied for treating the datasets to ensure statistical confidentiality and prevent unauthorised disclosure.	Describe the procedures that are used in protecting confidentiality.

Back to Brief Concepts

Institutional Mandate

3	Institutional mandate	Law, set of rules or other formal set of instructions assigning responsibility as well as the authority to an organisation for the collection, processing, and dissemination of statistics	
3.1	Legal acts and other agreements	Legal acts or other formal or informal agreements that assign responsibility as well as the authority to an agency for the collection, processing, and dissemination of statistics	State the national legal acts and/or other reporting agreements
3.2	Data sharing	Arrangements or procedures for data sharing and coordination between data producing agencies.	Describe the arrangements, procedures or agreements to facilitate data sharing and exchange between data producing agencies within the national statistical system
3.3	Release policy	Rules for disseminating statistical data to all interested parties	State if the release of the products is governed by some policy etc.
3.4	Release calendar	The schedule of statistical release dates.	State whether there is a release calendar for the statistical outputs from the process being reported, and if so, whether this calendar is publicly accessible and if yes, give a link or reference.
3.5	Frequency of dissemination	The time interval at which the statistics are disseminated over a given time period.	State the frequency with which the data are disseminated, e.g. monthly, quarterly, yearly.
3.6	Data access	The conditions and modalities by which users can access, use and interpret data	State the conditions and link on website from where the user can access the data For easy access of users, following details should also be mentioned about the dataset: **Title:** Name by which the data is known **Dataset Edition:** Edition of data (ex: first, second, final etc) **Dataset Reference data type:** Type of data entered in the field (ex: .txt, .dbf, .xls) **Presentation Format:** Presentation format of the data (ex: document, map, table, etc.) **Dataset Language:** language of any text in the data **Status/Version:** How updated is the data?

Back to Brief Concepts

Quality Management

4.	**Quality management**	**Systems and frameworks in place within an organisation to manage the quality of statistical products and processes.**	
4.1	Documentation on methodology	Descriptive text and references to methodological documents available.	List reference metadata files, methodological papers, summary documents and handbooks relevant to the statistical process.
4.2	Quality documentation	Documentation on procedures applied for quality management and quality assessment.	List relevant quality related documents, for example, other quality reports, studies.
4.3	Quality assurance	All systematic activities implemented that can be demonstrated to provide confidence that the processes will fulfil the requirements for the statistical output.	Describe the procedures (such as use of a general quality management system based on ISO 9000 series) to promote general quality management principles in the organisation. Describe the quality assurance framework used to implement statistical quality principles. Describe the quality assurance procedures specifically applied to the statistical process for which the report is being prepared, for example training courses, process monitoring, benchmarking, assessments, and use of best practices. Describe any ongoing or planned improvements in quality assurance procedures.
4.4	Quality assessment	Overall assessment of data quality, based on standard quality criteria.	Summarise the results of the most recent quality assessments and cross reference to the chapters in the report where the results are presented in more detail.

Accuracy and Reliability

Back to Brief Concepts

5	**Accuracy and reliability**	**Accuracy of data is the closeness of computations or estimates to the exact or true values that the statistics were intended to measure. Reliability of the data, defined as the closeness of the initial estimated value to the subsequent estimated value.**	
5.1	Sampling error	That part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a subset of the population is enumerated.	If probability sampling is used: • for user reports-provide the range of variation of the A1 indicator among key variables at user report level of detail; • for producer reports-provide the range of variation of the A1 indicator among key variables at producer report level of detail; • indicate the impact of sampling error on the overall accuracy of the results; • state how the calculation of sampling error is affected by adjustments for nonresponse, misclassifications and other sources of uncertainty, such as outlier treatment. If non-probability sampling is used: provide an assessment of representativeness, a motivation for the invoked model for estimation and risk of sampling bias

Back to Brief Concepts

Timeliness and Punctuality

6	**Timeliness**	**The timeliness of the data collection release to be compiled.**	
6.1	Timeliness	Length of time between data availability, the event or phenomenon the data describe, and final release to its users.	Outline the reasons for the time lag, if any. Outline efforts to reduce time lag in future.

Coherence and Comparability

Back to Brief Concepts

7	**Coherence and comparability**	**Adequacy of statistics to be reliably combined in different ways and for various uses and the extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics**	
7.1	Comparability – over time	The extent to which statistics are comparable or reconcilable over time	Provide information on possible limitations in the use of data for comparisons over time. Distinguish three broad possibilities: 1. There have been no changes, in which case this should be reported. 2. There have been some changes but not enough to warrant the designation of a break in series. 3. There have been sufficient changes to warrant the designation of a break in series.
7.6	Coherence	The extent to which statistics are reconcilable with System of National Accounts.	For producer reports only. Where relevant, the results of comparisons with the System of National Account framework / Other Statistical Standards and feedback from System of National Accounts / Other Statistical Standards with respect to coherence and accuracy problems should be reported and should be a trigger for further investigation.

Statistical Processing

Back to Brief Concepts

8	Statistical processing	Any statistical processing undertaken to finalise the data	
8.1	Source data type	Characteristics and components of the raw statistical data used for compiling statistical aggregates.	Indicate if the data are based on a survey, administrative data, multiple data sources, or macro-aggregates. In the event of multisource or macro-aggregate processes describe each source. For each dataset from an administrative source, summarise the source, its primary purpose, and the most important data items acquired
8.2	Frequency of data collection	Time interval at which the source data are collected	Indicate the frequency of data collection (e.g. monthly, quarterly, annually, or continuous).
8.3	Data collection method	Method applied for gathering data for official statistics.	For each source of survey data: • describe the method(s) used to gather data from respondents; • annex or hyperlink the questionnaires(s). For each source of administrative data: • describe the acquisition process and how it was tested. For all sources: • describe the types of checks applied at the time of data entry.

Statistical Processing (Contd…)

Back to Brief Concepts

8.4	Data validation	Process of monitoring the results of data compilation and ensuring the quality of statistical results	Describe the procedures for checking and validating the source data and how the results are monitored and used. Describe the procedures for validating the aggregate output data (statistics) after compilation, including checking coverage and response rates, and comparing with data for previous cycles and with expectations. List other output datasets to which the data relate and outline the procedures for identifying inconsistencies between the output data and these other datasets
8.5	Data compilation	Operations performed on data to derive new information according to a given set of rules.	Describe the procedures for imputation, the most common reasons for imputation and imputation rates within each of the main strata. Describe the likely impact of imputation. Describe the procedures to derive new variables and to calculate aggregates and complex statistics. Describe the procedures for adjustment for non-response and the corrections to the design weights to account for differences in response rates. Describe the calculation of design weights, including calibration (if used). Describe the procedures for combining input data from different sources. Provide the ratio of the number of replaced values to the total number of values for a given variable. Specific reference to formula shall be made. The formula or mathematical equation used while computing different variables in the report may be described here in a structured format showing the Numerator; Denominator and Multiplier used for computing the same

Metadata Update

Back to Brief Concepts

9	Metadata update	The date on which the metadata element was inserted or modified in the database.	
9.1	Metadata last posted	Date of the latest dissemination of the metadata	The date when the complete set of metadata was last disseminated as a block should be provided (manually, or automatically by the metadata system).
9.2	Metadata last update	Date of last update of the content of the metadata.	The date when any metadata were last updated should be provided (manually, or automatically by the metadata system).

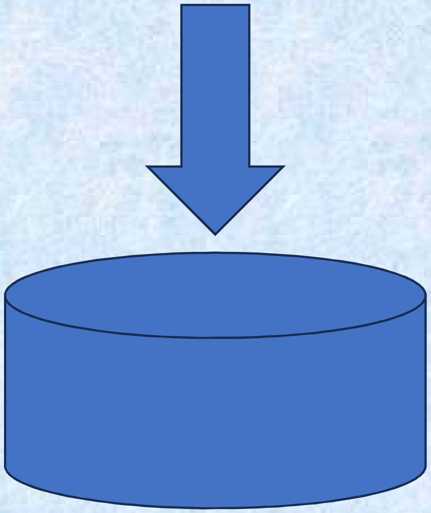
Metadata Structure for Statistical Products as Available in Public Domain

Use Cases of Metadata Compliance of MoSPI

What Next

- From ”paper” to database

- Information in the word document



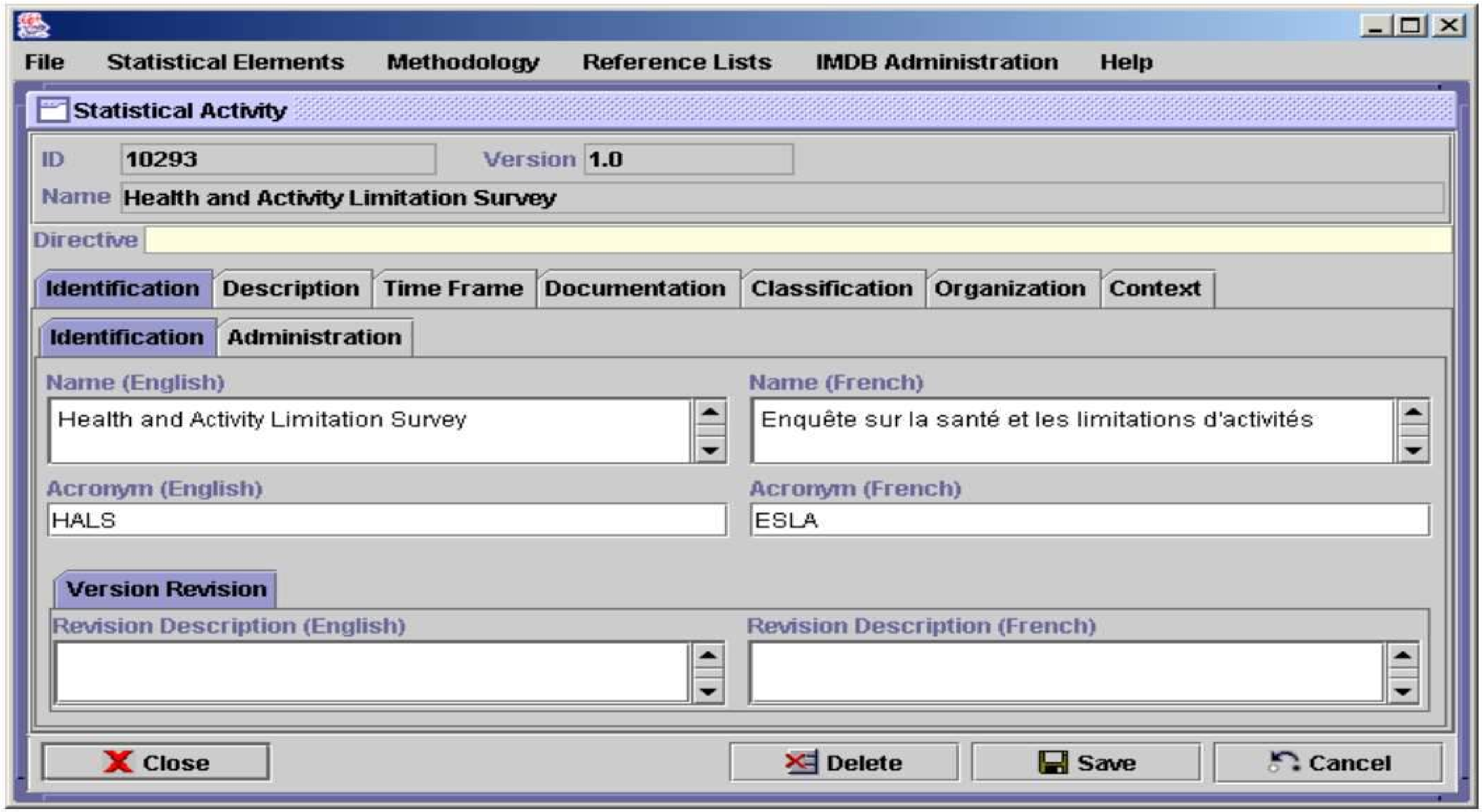
Revised NMDS

Introduction of Two new Concepts

- **2.3 International/National Standards Classification etc.**
- **8.6 Identifier of a statistical indicator**

➡ NMDS

Loading metadata into India Integrated Metadata Base (IIMDB)



File
Statistical Elements
Methodology
Reference Lists
IMDB Administration
Help
Statistical Activity
Identification
Description
Time Frame
Documentation
Classification
Organization
Context
Identification
Administration
Version Revision
Close
Delete
Save
Cancel

ID: 10293
Version: 1.0
Name: Health and Activity Limitation Survey
Directive:

Name (English): Health and Activity Limitation Survey
Name (French): Enquête sur la santé et les limitations d'activités
Acronym (English): HALS
Acronym (French): ESLA
Revision Description (English):
Revision Description (French):

Thank You

GOVERNMENT OF INDIA
MINISTRY OF STATISTICS AND
PROGRAMME IMPLEMENTATION

NSO INDIA
Data for Development

Data Dissemination: IT Initiatives of MoSPI

28th Conference of Central and State Statistical Organizations

Session XI,13th August 2024

Rohit Bhardwaj,DDG

Our Endeavour

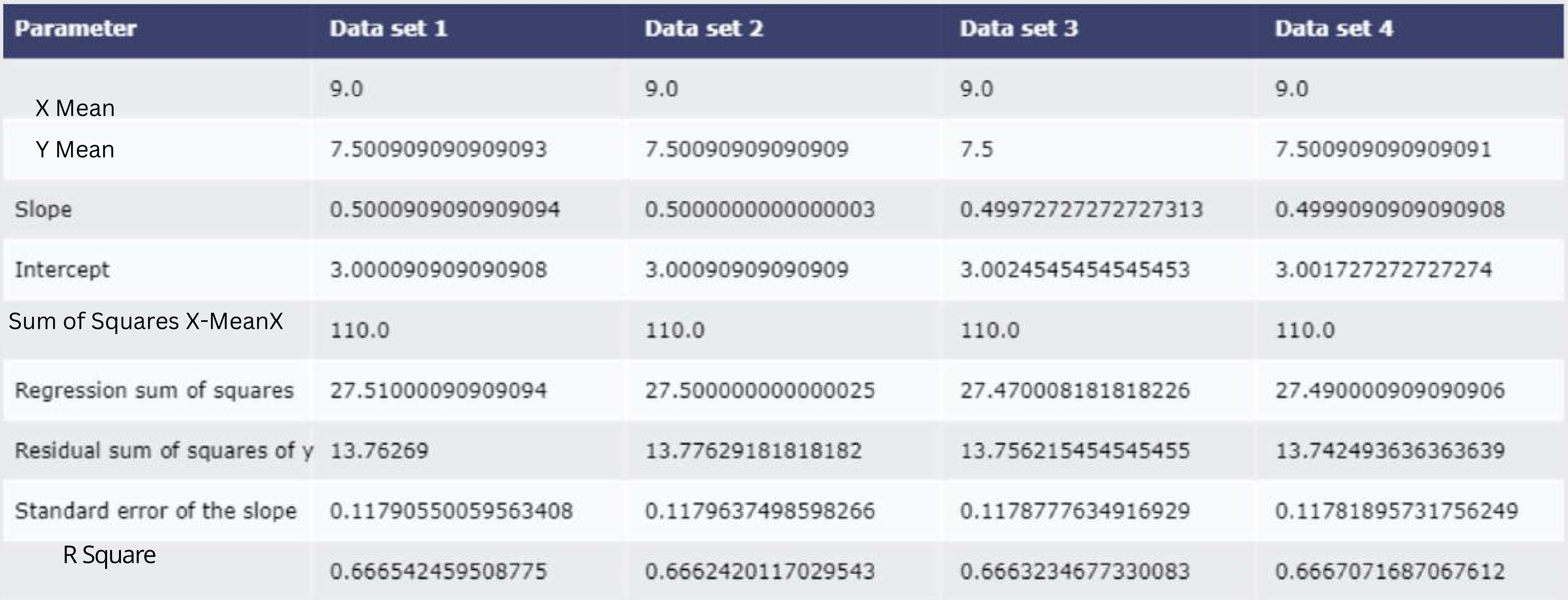
Data Visualizations
for better data understanding

eSankhyiki portal
for better data access & management

Data visualizations: the Need

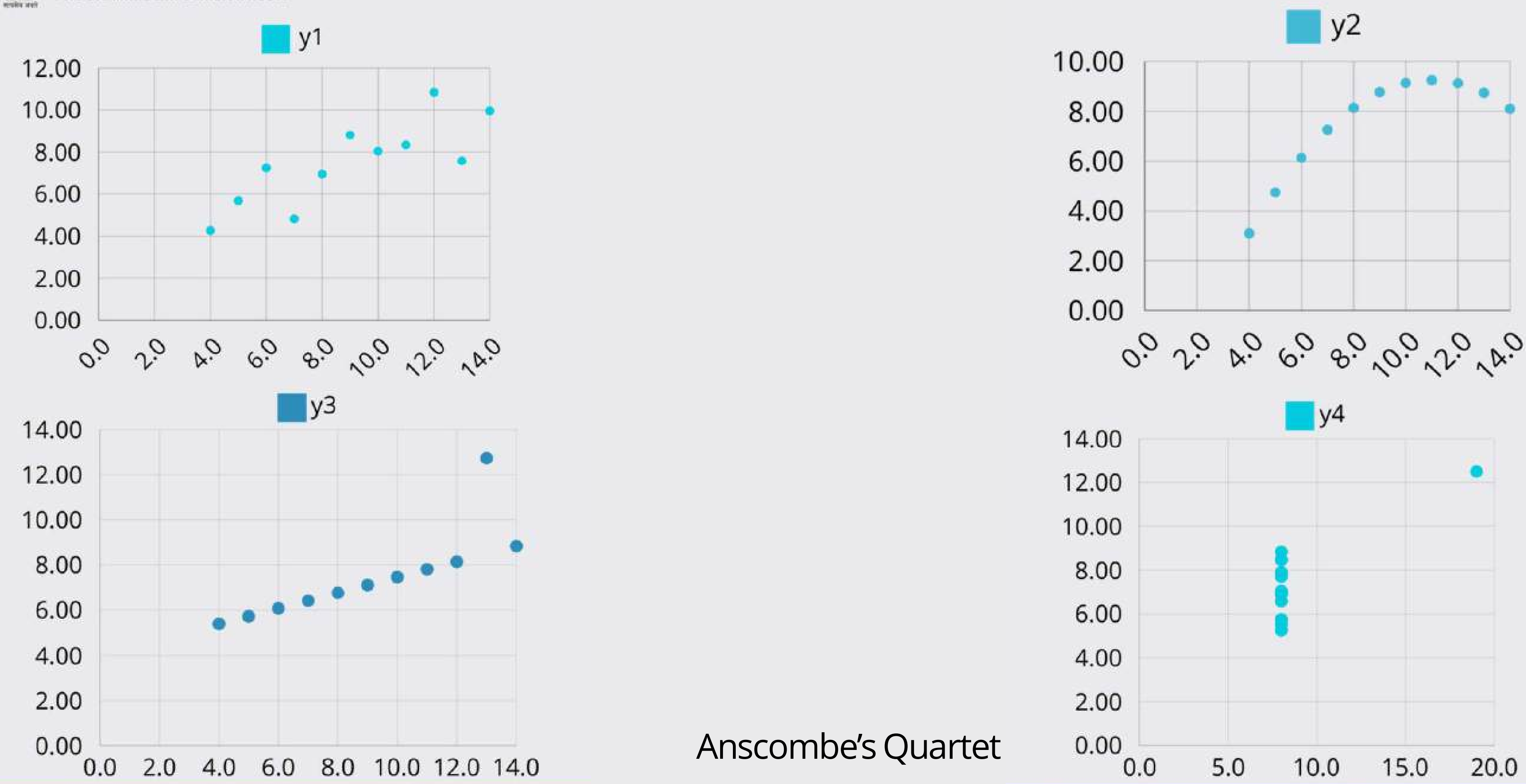
x123	y1	y2	y3	x4	y4
10	8.04	9.14	7.46	8	6.58
8	6.95	8.14	6.77	8	5.76
13	7.58	8.74	12.74	8	7.71
9	8.81	8.77	7.11	8	8.84
11	8.33	9.26	7.81	8	8.47
14	9.96	8.1	8.84	8	7.04
6	7.24	6.13	6.08	8	5.25
4	4.26	3.1	5.39	19	12.5
12	10.84	9.13	8.15	8	5.56
7	4.82	7.26	6.42	8	7.91
5	5.68	4.74	5.73	8	6.89

Data visualizations: the Need



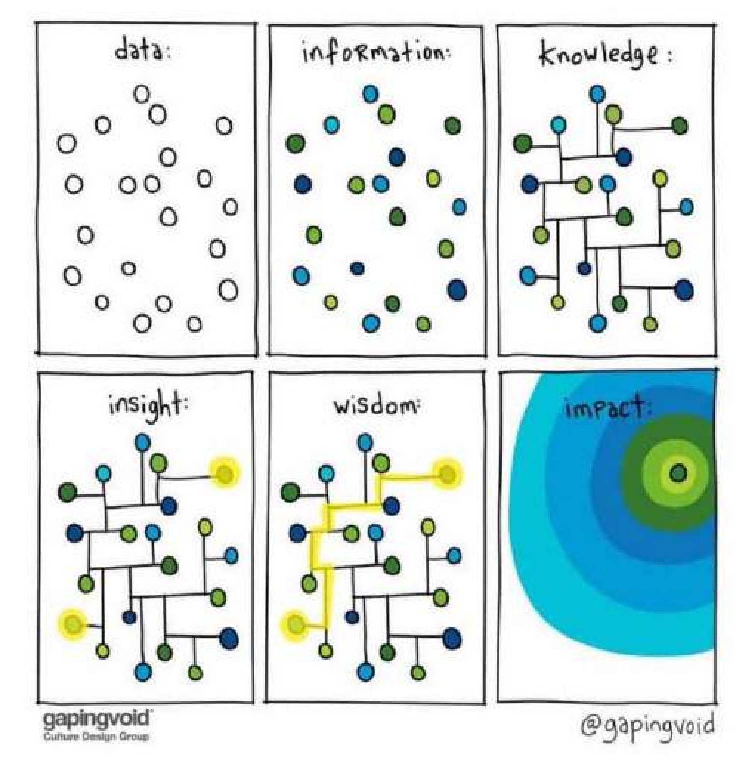
Parameter	Data set 1	Data set 2	Data set 3	Data set 4
X Mean	9.0	9.0	9.0	9.0
Y Mean	7.500909090909093	7.50090909090909	7.5	7.500909090909091
Slope	0.5000909090909094	0.5000000000000003	0.49972727272727313	0.4999090909090908
Intercept	3.000090909090908	3.00090909090909	3.0024545454545453	3.001727272727274
Sum of Squares X-MeanX	110.0	110.0	110.0	110.0
Regression sum of squares	27.51000090909094	27.500000000000025	27.470008181818226	27.490000909090906
Residual sum of squares of y	13.76269	13.77629181818182	13.756215454545455	13.742493636363639
Standard error of the slope	0.11790550059563408	0.1179637498598266	0.1178777634916929	0.11781895731756249
R Square	0.666542459508775	0.6662420117029543	0.6663234677330083	0.6667071687067612

Data visualizations: Look Before You Leap



Anscombe's Quartet

Data visualizations: the Advantages


data:
information:
knowledge:
insight:
wisdom:
impact:
gapingvoid
Culture Design Group
@gapingvoid

- Ease of data understanding
- User empowerment
- Adds Value to data
- To enhance statistical Analysis

Data visualizations: the Approach

- Follow global best practices
  - The Chartmaker directory
  - datavizcatalogue.com/
  - OurWorldinData
  - World Bank Data Bank
  - FRED Web site

- Create in-house capacity of ISS and SSS officers for sustainability

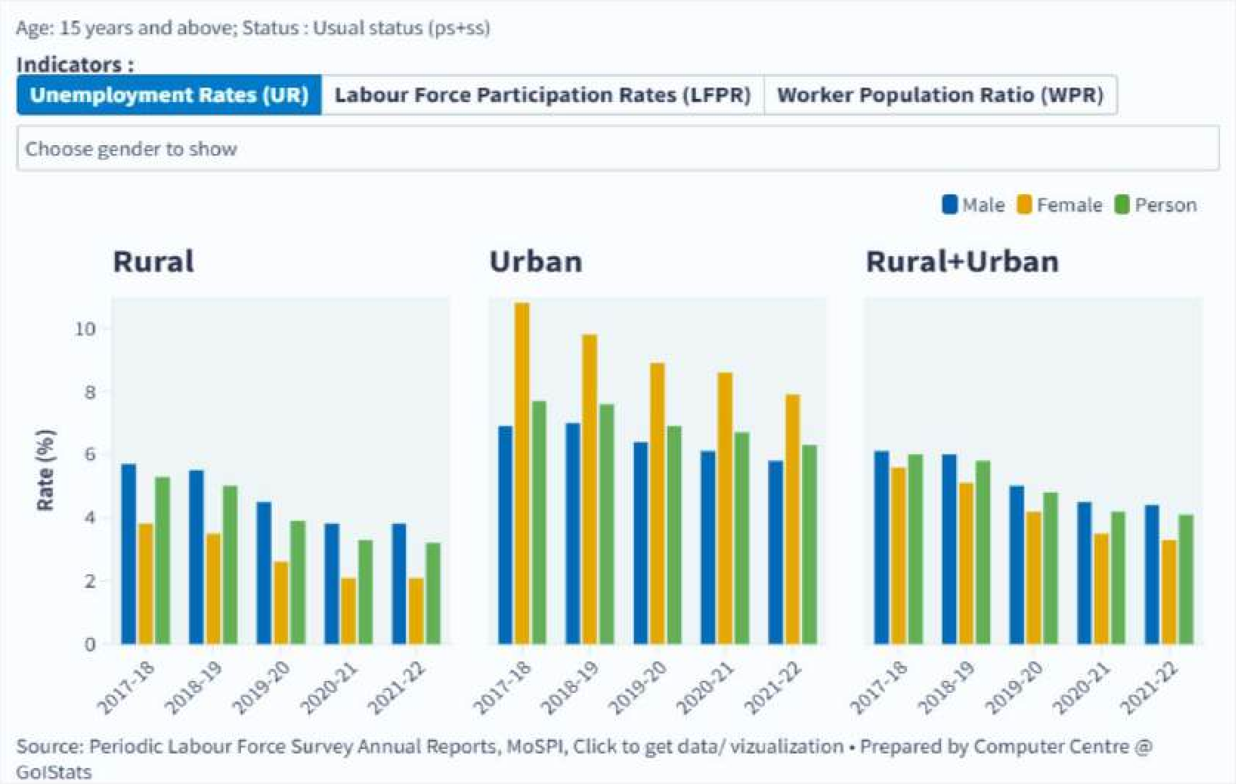
- Use tools such as Flourish, Datawrapper, Tableau, Dash etc.

- Create in continuum

Data Visualizations Features

- Separate DataViz section on MoSPI website
- Consists of 1400+ visualizations
- Categorized as per themes- NAS, CPI, IIP, ASI, PLFS, HCES, MIS Covered
- Filter and Search options
- Tree map, sunburst, streamgraph, radial tree, bubble chart, heatmap and many more...

Employment Indicators in India

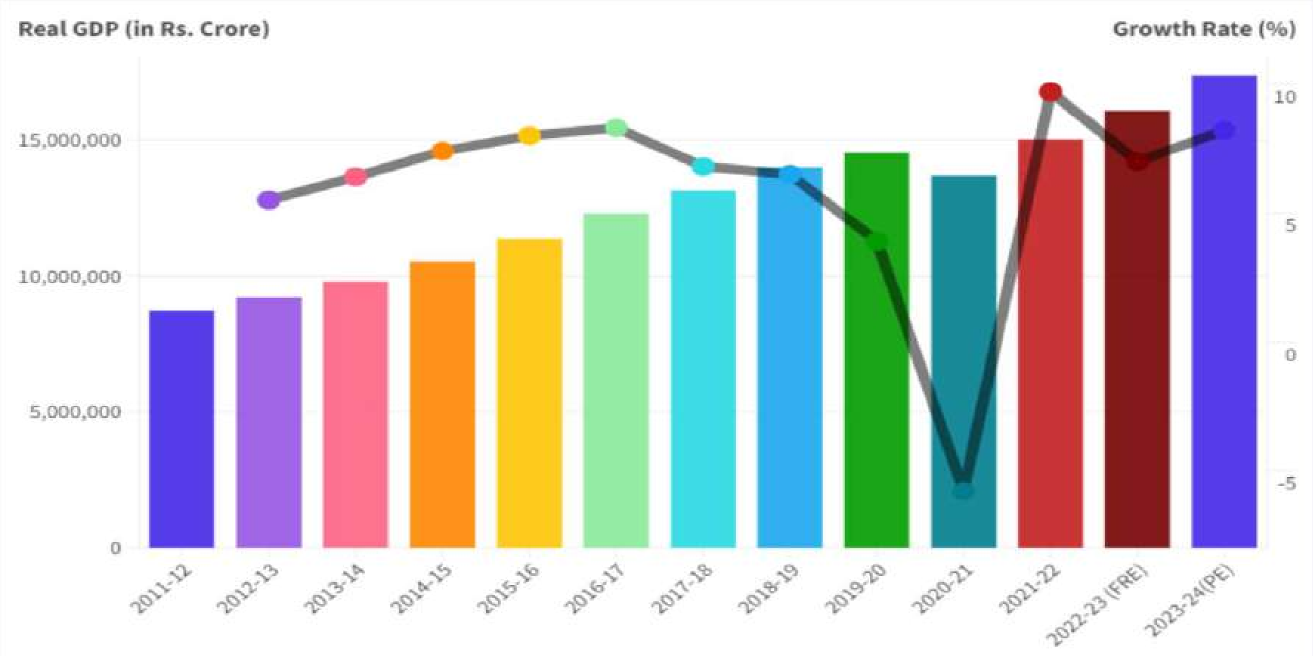


Age: 15 years and above; Status : Usual status (ps+ss)

Indicators :

Source: Periodic Labour Force Survey Annual Reports, MoSPI, Click to get data/ vizualization • Prepared by Computer Centre @ GoIStats

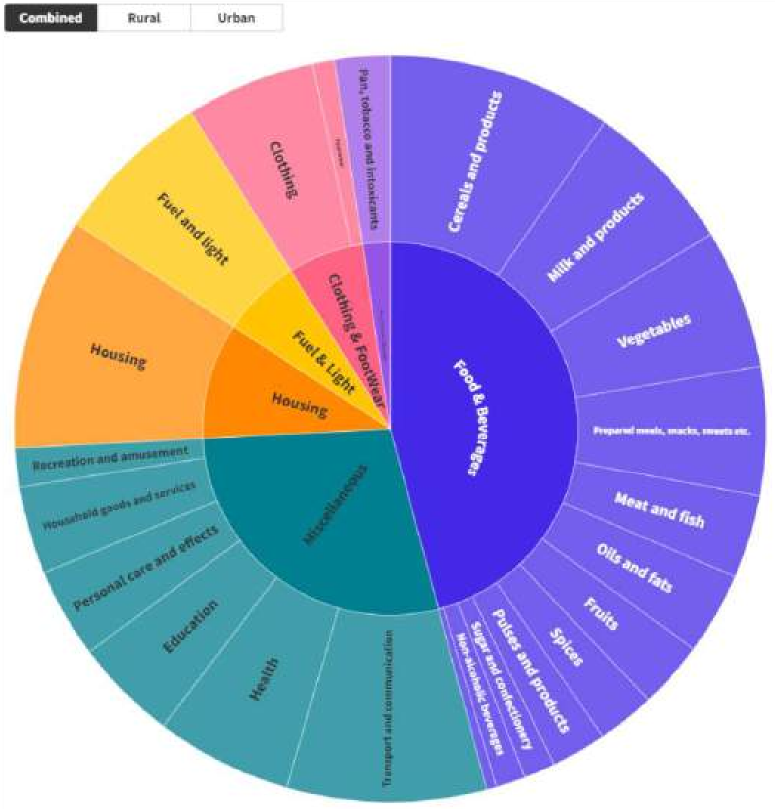
Annual GDP Estimates (in Rs. Crore) and Growth Rates (%) at Constant Prices (Base 2011–12)



Source: MoSPI, India (Publication: PROVISIONAL ESTIMATES OF ANNUAL GDP FOR 2023-24), Click to get data/ vizualization • Prepared by Computer Centre @GoIStats

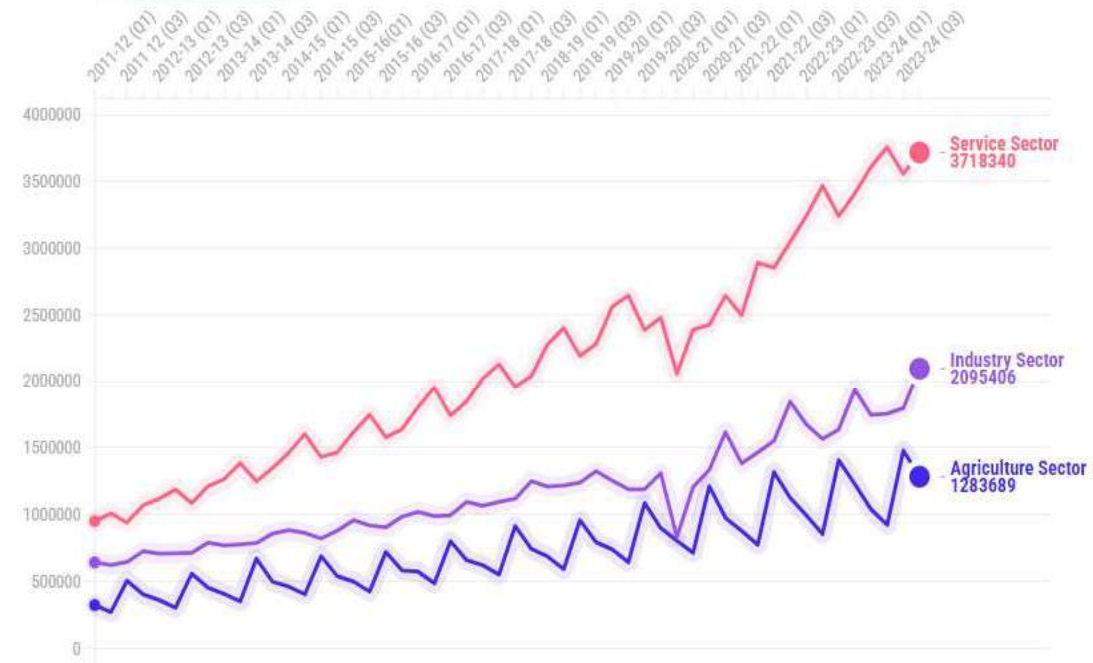
NNote: Growth Rates Calculated with respect to Previous Year; FRE: First Revise Estimate; PE: Provisional Estimates

Percentage Share of Different Groups of items in the basket of CPI (Base 2012)



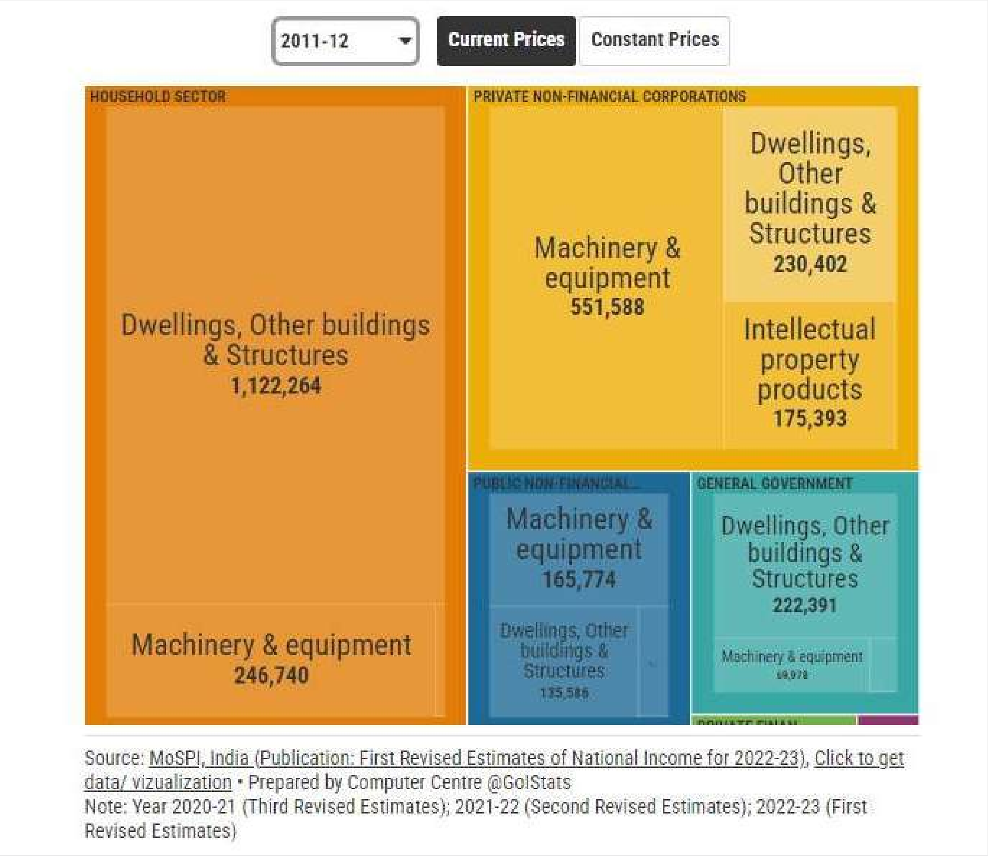
Source: MoSPI India, Click to get data/ vizualization • prepared by Computer Centre @ GoIStats

Quarterly Gross Value Added by Different Economic Sectors (in Rs. Crore) (Current/Constant Prices) (Base Year: 2011-12)



Source: MoSPI, India (Publication: PROVISIONAL ESTIMATES OF ANNUAL GDP FOR 2023-24 AND QUARTERLY ESTIMATES OF GDP FOR Q4 OF 2023-24), Click to get data/ vizualization • Prepared by Computer Centre @ GoIStats
Note: The above classification (Agriculture, Industry and Services Sector) is as per Special Data Dissemination System (SDDS)

Gross Fixed Capital Formation by asset & institutional sector (in Rs. Crore) (Current/ Constant Prices)
(Base Year: 2011-12)



Source: MoSPI, India (Publication: First Revised Estimates of National Income for 2022-23), Click to get data/ vizualization • Prepared by Computer Centre @GoIStats
Note: Year 2020-21 (Third Revised Estimates); 2021-22 (Second Revised Estimates); 2022-23 (First Revised Estimates)

eSankhyiki portal

A comprehensive data management & sharing system for dissemination of official statistics in the country

Web- based

Interactive

Database

Data Dissemination Practice

MoSPI Website used for dissemination of Macro Data

Disparate Data Sets

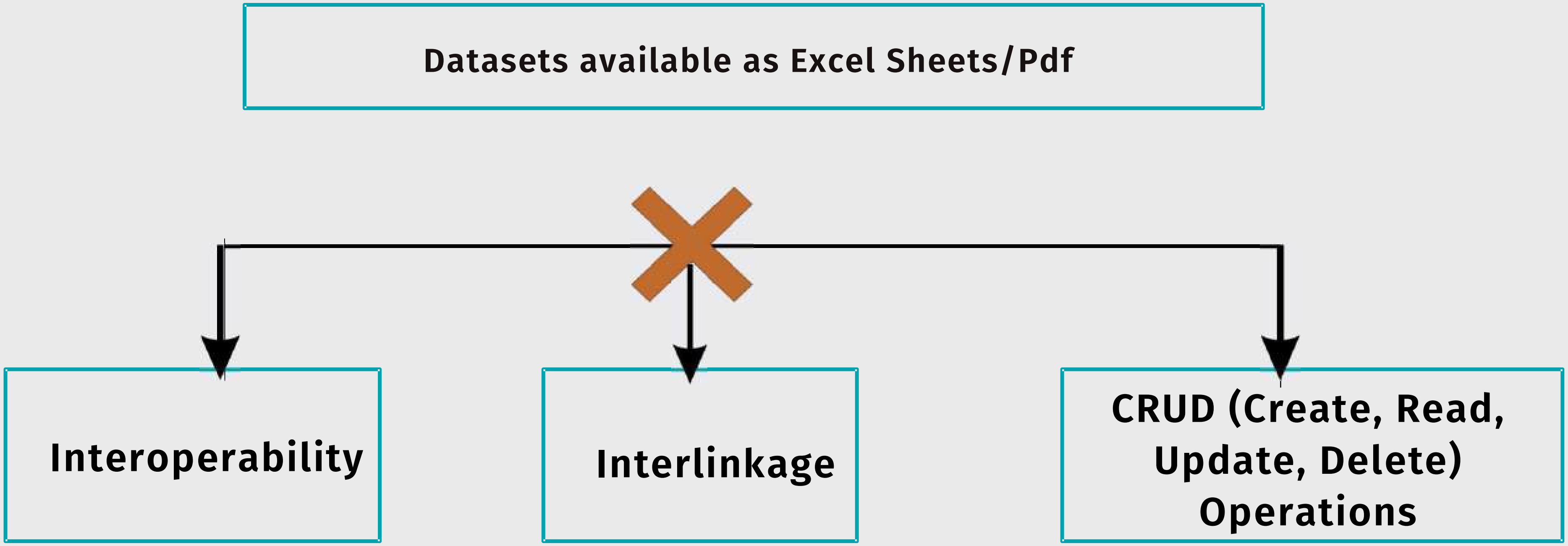
41000 tables in 1100 documents

Disseminated as files

Either as Excel sheet or pdf

Filter/ Search not available within data sets

Bottleneck in Realising True Potential of Data



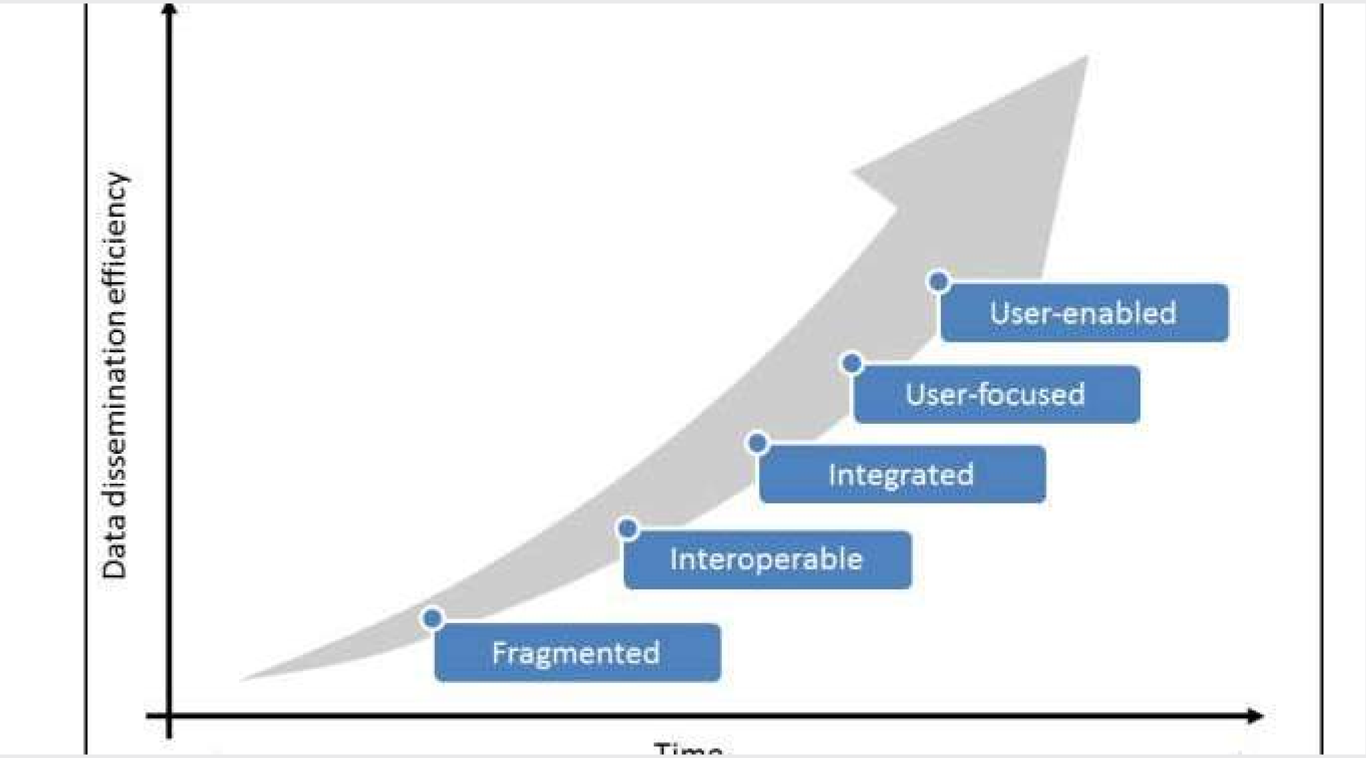
Datasets available as Excel Sheets/Pdf

Interoperability

Interlinkage

CRUD (Create, Read, Update, Delete) Operations

Data Portal maturity model


Data dissemination efficiency
User-enabled
User-focused
Integrated
Interoperable
Fragmented
Time

Source – Paris 21 Discussion Paper no:8 ; April 2016

Portals world over

Increasing focus on analytics →

OGD

Diverse range of data sets made accessible to public

World Bank

Open Data Catalogue : Listing of available world bank datasets

‘*Databank*’ is the Analysis and visualization tool that contain collection of time series data

NSO

Focus on dissemination of official statistics data (macro data micro data , metadata etc.)

Analytics provision includes Data Visualization, Custom datasets etc. along with creating data products

BIS

‘*Statistical warehouse*‘ is the Data Dissemination and Analytical tool which includes the financial and monetary data to produce research reports and statistics

Central Bank

Provide access to financial and economic data including historical data and time series

Analytics provision includes Data Visualization, Data downloads, API access to build custom application

IMF

Provide access to financial and economic databases which provides projections on global economy

Analytics provision includes Data Visualization, Data downloads, API access to build custom application

← Increasing focus on data accessibility

Background Work

Desk Research	○ UNSD, PARIS21 , Dr Rangarajan Committee report/ Publications etc.
Meetings and Consultations	○ With various expert institutions/ stakeholders viz. World Bank, Federal Reserve Economic Data (US), MeitY, RBI etc.
Report collation	○ Metadata details across various ministries/ departments ○ Report on alternate data sources by NSC
Examination of existing portals	○ Multilateral organizations, Central Reserve banks, NSO and Open data portals of various countries, including India.

Key Takeaways

Focus on data access and management

Unified portal a necessity to lessen burden on user to explore multiple sites for data

Lend value to disseminated data by building trust

eSankhyiki portal Launched on 29th June,2024

Scan to access eSankhyiki

eSankhyiki portal

Macro Indicators

Searchable time series data of 4 products consisting of more than 27 Lakh records with custom download, visualization and API facility

Data Catalogue

More than 2300 data assets of 7 MoSPI products with facility for search, preview, visualization and download along with metadata.

Data Catalogue

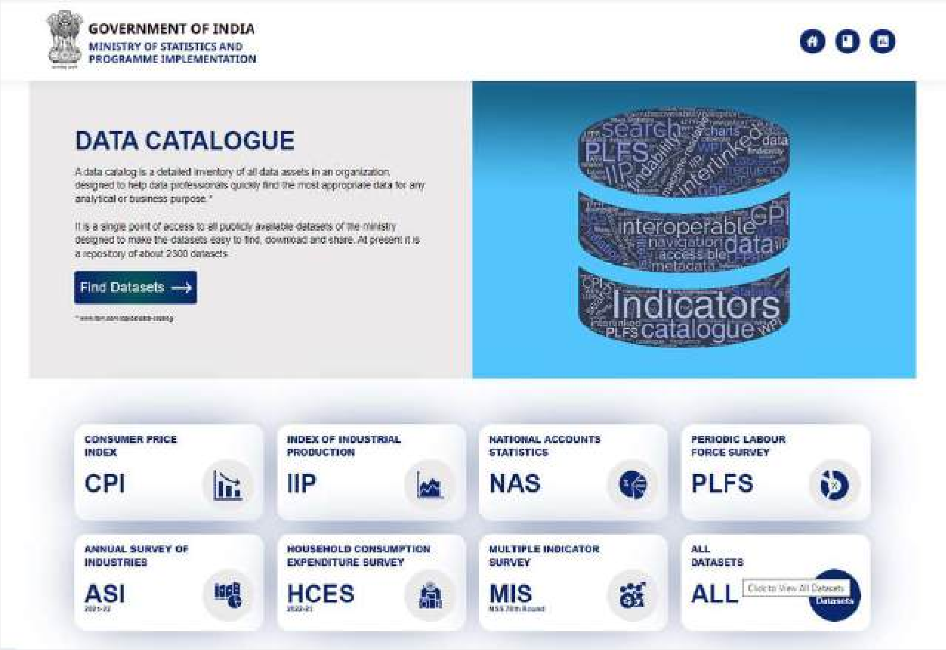
GOVERNMENT OF INDIA
MINISTRY OF STATISTICS AND
PROGRAMME IMPLEMENTATION

Data Catalogue

Listing of data assets of key MoSPI indicators

Intuitive navigation with **search**, **filtering** and **download** options

Dataset specific metadata, visualizations and searchable tables


GOVERNMENT OF INDIA
MINISTRY OF STATISTICS AND
PROGRAMME IMPLEMENTATION
DATA CATALOGUE
Find Datasets →
search
PLFS
IIP
interlinked
data
interoperable
CPI
navigation
accessible
metadata
Indicators
catalogue
CONSUMER PRICE INDEX
CPI
INDEX OF INDUSTRIAL PRODUCTION
IIP
NATIONAL ACCOUNTS STATISTICS
NAS
PERIODIC LABOUR FORCE SURVEY
PLFS
ANNUAL SURVEY OF INDUSTRIES
ASI
2021-22
HOUSEHOLD CONSUMPTION EXPENDITURE SURVEY
HCES
2022-23
MULTIPLE INDICATOR SURVEY
MIS
ALL DATASETS
ALL
Click to View All Datasets
Datasets

A data catalog is a detailed inventory of all data assets in an organization, designed to help data professionals quickly find the most appropriate data for any analytical or business purpose.*

It is a single point of access to all publicly available datasets of the ministry designed to make the datasets easy to find, download and share. At present it is a repository of about 2300 datasets.

Data Catalog Features

1

Constitutes all tables of key products in a single place

2

Search tables as per keywords/ table no/ unique table title in Data Catalog

3

Access to metadata pertaining to each table displayed alongside

4

Downloadable data in MS Excel formats

5

Access to data/ visualization/ preview table in Single click

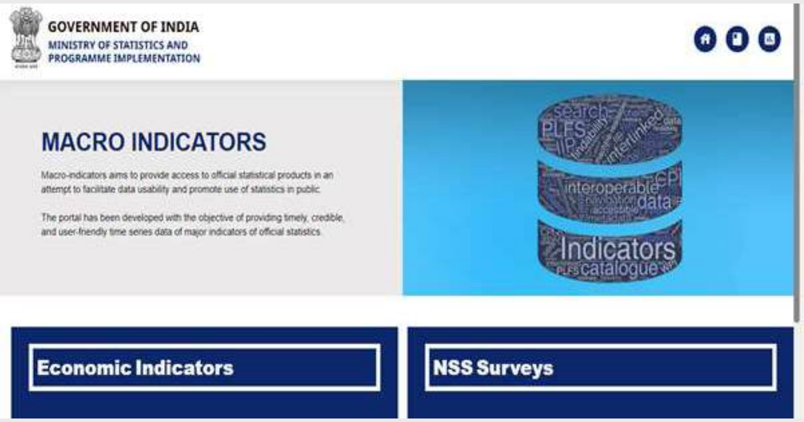
6

Options to filter data

Macro Indicators

Macro Indicators

- Time series data for 10 years of Key indicators
- Intuitive navigation with **search**, **filtering** and **download** options
- Metadata, visualizations and API
- Mobile Compliant



GOVERNMENT OF INDIA
MINISTRY OF STATISTICS AND
PROGRAMME IMPLEMENTATION

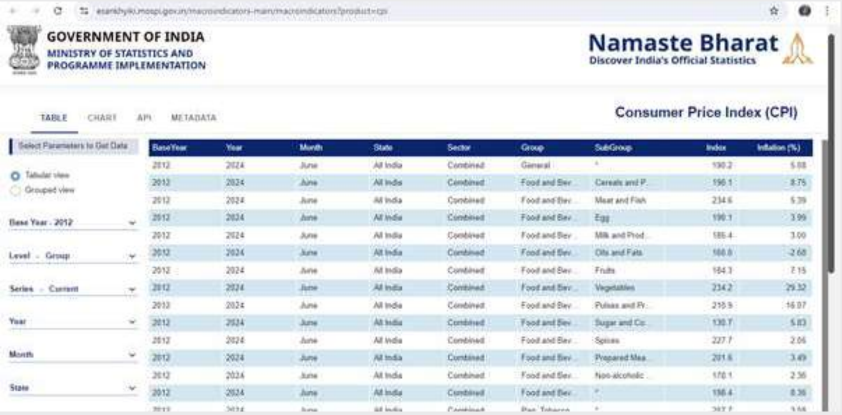
MACRO INDICATORS

Macro-indicators aims to provide access to official statistical products in an attempt to facilitate data usability and promote use of statistics in public.

The portal has been developed with the objective of providing timely, credible, and user-friendly time series data of major indicators of official statistics.

Economic Indicators

NSS Surveys



esankhyiki.mospi.gov.in/macroindicators-main/macroindicators?product=cpi

GOVERNMENT OF INDIA
MINISTRY OF STATISTICS AND
PROGRAMME IMPLEMENTATION

Namaste Bharat
Discover India's Official Statistics

Consumer Price Index (CPI)

Select Parameters to Get Data

● Tabular view
○ Grouped view

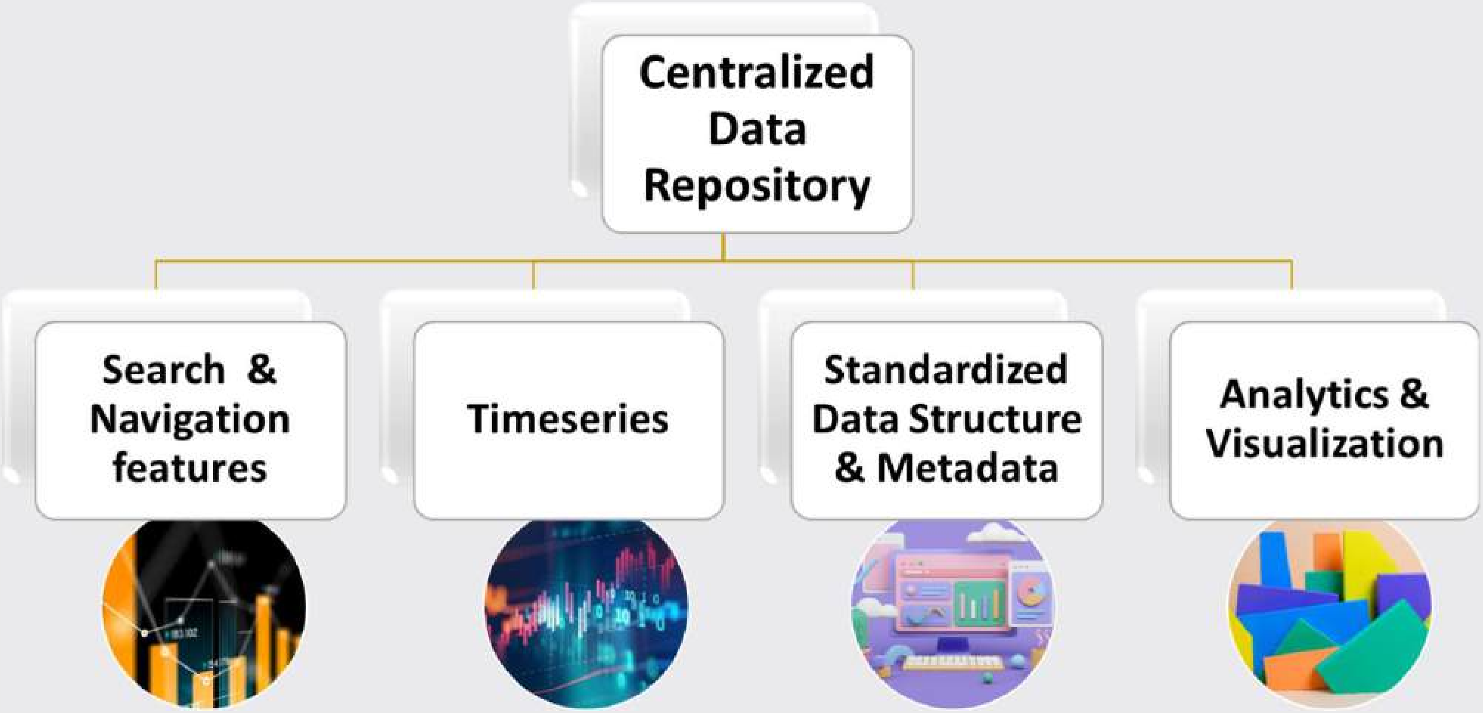
Base Year - 2012
Level - Group
Series - Current
Year
Month
State

BaseYear	Year	Month	State	Sector	Group	SubGroup	Index	Inflation (%)
2012	2024	June	All India	Combined	General	*	190.2	5.08
2012	2024	June	All India	Combined	Food and Bev...	Cereals and P...	196.1	8.75
2012	2024	June	All India	Combined	Food and Bev...	Meat and Fish	234.6	5.39
2012	2024	June	All India	Combined	Food and Bev...	Egg	190.1	3.99
2012	2024	June	All India	Combined	Food and Bev...	Milk and Prod...	186.4	3.00
2012	2024	June	All India	Combined	Food and Bev...	Oils and Fats	168.8	-2.68
2012	2024	June	All India	Combined	Food and Bev...	Fruits	184.3	7.15
2012	2024	June	All India	Combined	Food and Bev...	Vegetables	234.2	29.32
2012	2024	June	All India	Combined	Food and Bev...	Pulses and Pr...	215.9	16.07
2012	2024	June	All India	Combined	Food and Bev...	Sugar and Co...	130.7	5.83
2012	2024	June	All India	Combined	Food and Bev...	Spices	227.7	2.06
2012	2024	June	All India	Combined	Food and Bev...	Prepared Mea...	201.6	3.49
2012	2024	June	All India	Combined	Food and Bev...	Non-alcoholic ...	170.1	2.36
2012	2024	June	All India	Combined	Food and Bev...	*	198.4	8.36

Central Data Repository

Central Data Repository:

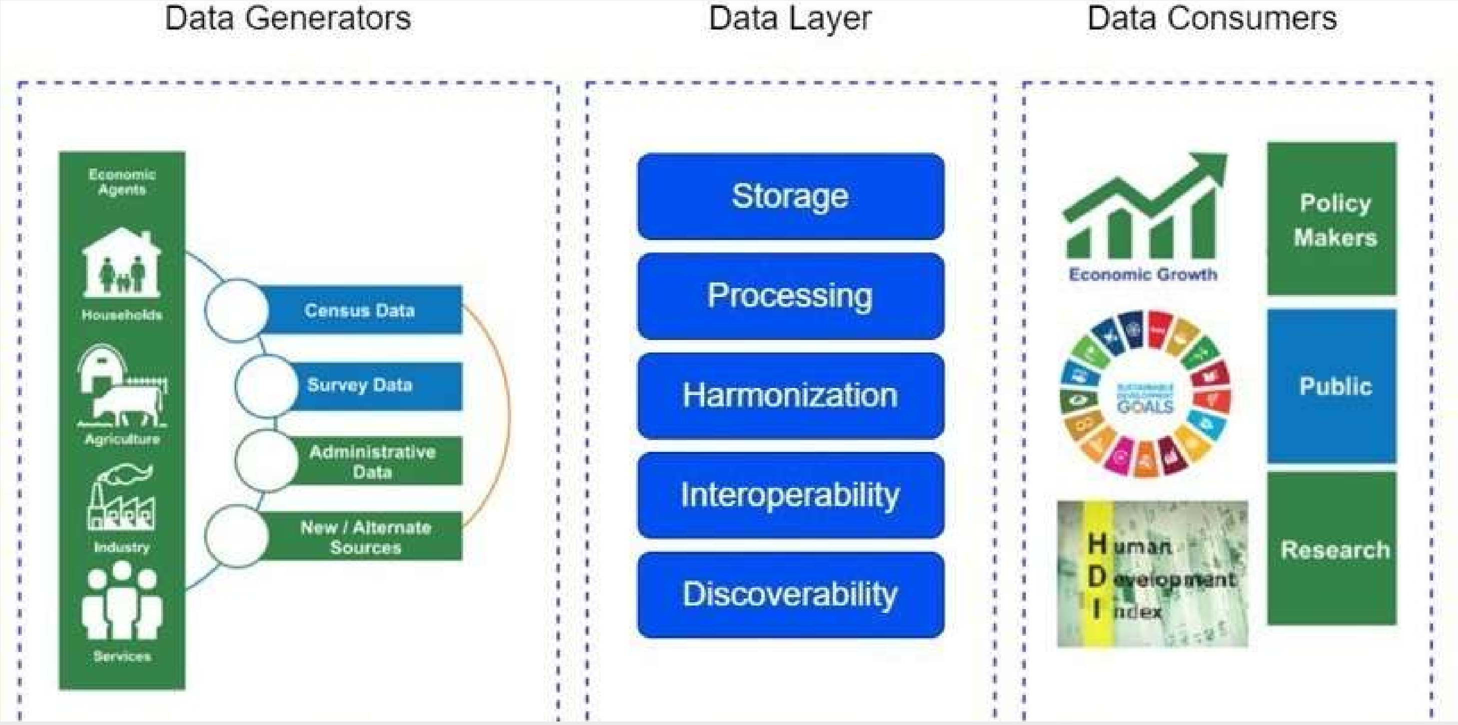
Based on Global Standards



Centralized Data Repository

- Search & Navigation features
- Timeseries
- Standardized Data Structure & Metadata
- Analytics & Visualization

Information architecture

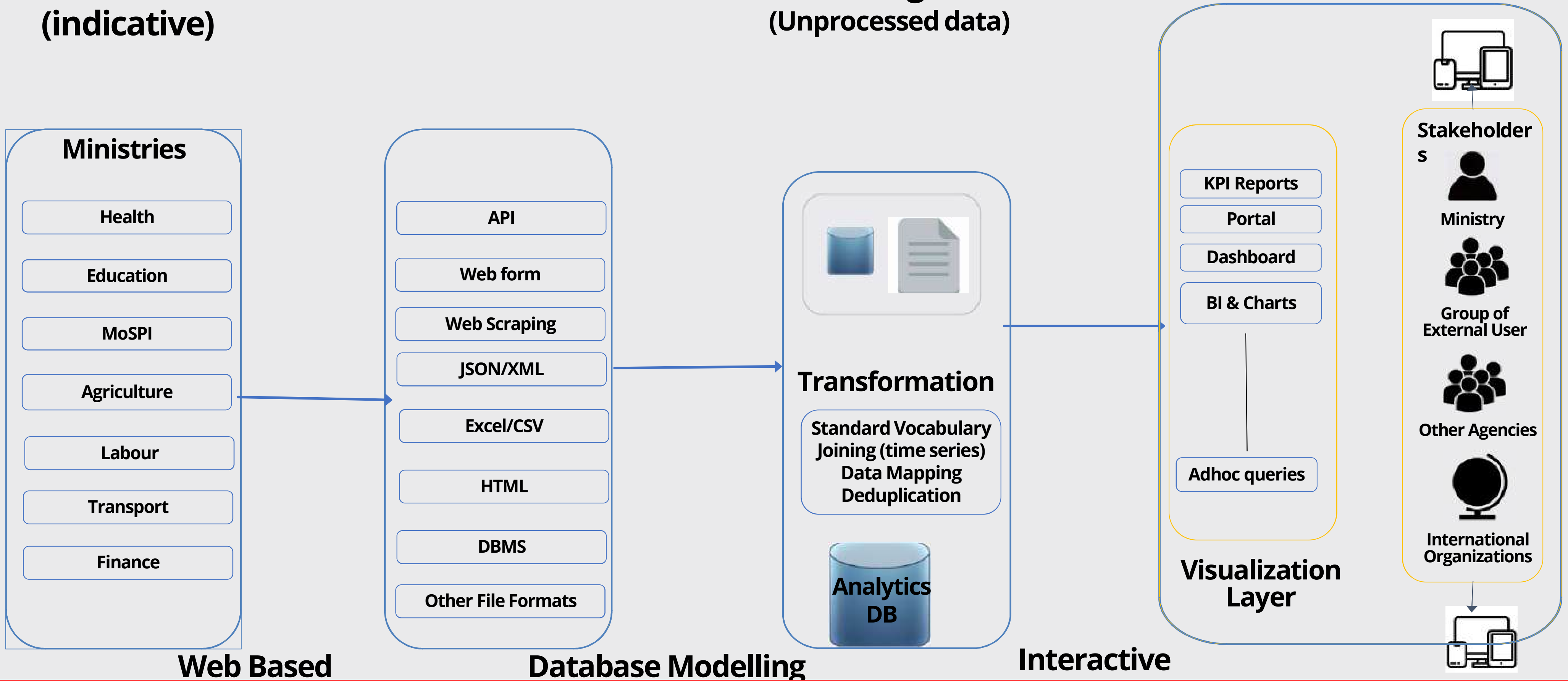

Data Generators
Data Layer
Data Consumers
Economic Agents
Households
Agriculture
Industry
Services
Census Data
Survey Data
Administrative Data
New / Alternate Sources
Storage
Processing
Harmonization
Interoperability
Discoverability
Economic Growth
SUSTAINABLE DEVELOPMENT GOALS
Human Development Index
Policy Makers
Public
Research

Data acquisition through **machine to machine exchange** (APIs) or any other mechanism with least manual intervention

ETL layer to transform and process the data into a standard structure to make it interoperable

Creation of time series and making data sharable through APIs and other formats

Technology Architecture



Data Sources (indicative)

Data ingestion

Data Storage
(Unprocessed data)

Data Dissemination

Ministries
- Health
- Education
- MoSPI
- Agriculture
- Labour
- Transport
- Finance

- API
- Web form
- Web Scraping
- JSON/XML
- Excel/CSV
- HTML
- DBMS
- Other File Formats

Transformation
- Standard Vocabulary
- Joining (time series)
- Data Mapping
- Deduplication

Analytics DB

- KPI Reports
- Portal
- Dashboard
- BI & Charts
- Adhoc queries

Visualization Layer

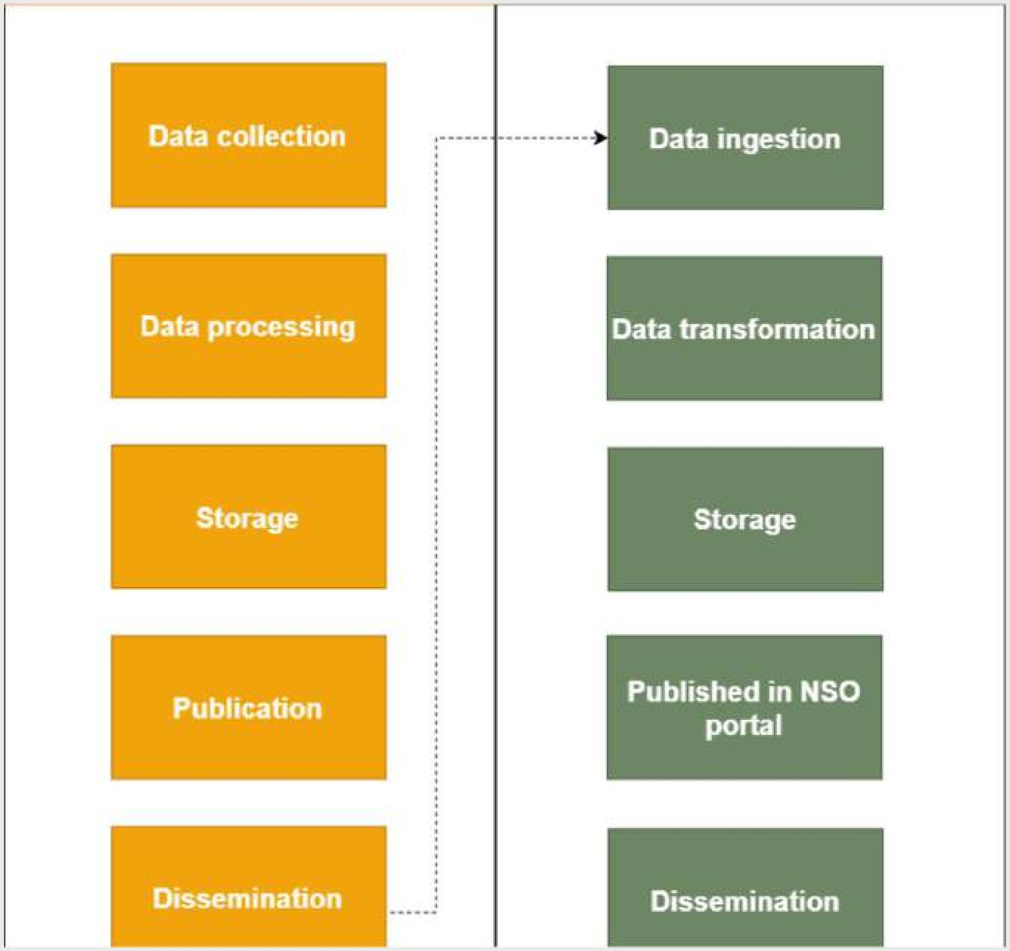
Stakeholders
- Ministry
- Group of External User
- Other Agencies
- International Organizations

Web Based

Database Modelling

Interactive

Indicative Data Flow



Data collection
Data processing
Storage
Publication
Dissemination
Data ingestion
Data transformation
Storage
Published in NSO portal
Dissemination

Portal Features

① Management of Data through standard data modelling

- Centralised data repository
- Data update on regular schedule
- Metadata accessibility
- Easy and Scalable Data Management
- Integration by APIs

② General required Features

- Data discovery in minimum clicks
- Collaboration and Sharing.
- User friendly search
- Comprehensive data coverage
- Data security and privacy
- User centric design

③ Analytics & Data Dissemination

- Filter data as per Analysis requiremet
- Time Series Data
- Interoperable, Time Series dissemination
- Advance analytics
- Data download in multiple formats
- Dissemination as per International requirements/ standards

④ Visualization

- Pre developed Visualization
- Visualization Builder
- Parameters for filtering/ customization
- Infographics

eSankhyiki portal

Outcomes

- Single Source of all Sharable data
- Dissemination with better data accessibility and management
- Enhanced User Experience
- Provide inputs for Policy making
- Availability of Data Analytics and Visualization tools
- Elevate trust in official data

Data are just not numbers, they speak volumes

-Prof P. C. Mahalanobis

Email: rohitb.rgi@nic.in

National Metadata Structure (NMDS)

A Standard Set of Guidelines for Ensuring Quality of Statistical Products Released

Objectives

- Promote reporting for each type of statistical process and its outputs across different Ministries/Divisions/Departments of NSO, hence facilitating comparisons across processes and outputs;

- Ensure that producer reports contain all the information required to facilitate identification of quality issues and potential improvements in statistical processes and their outputs; and

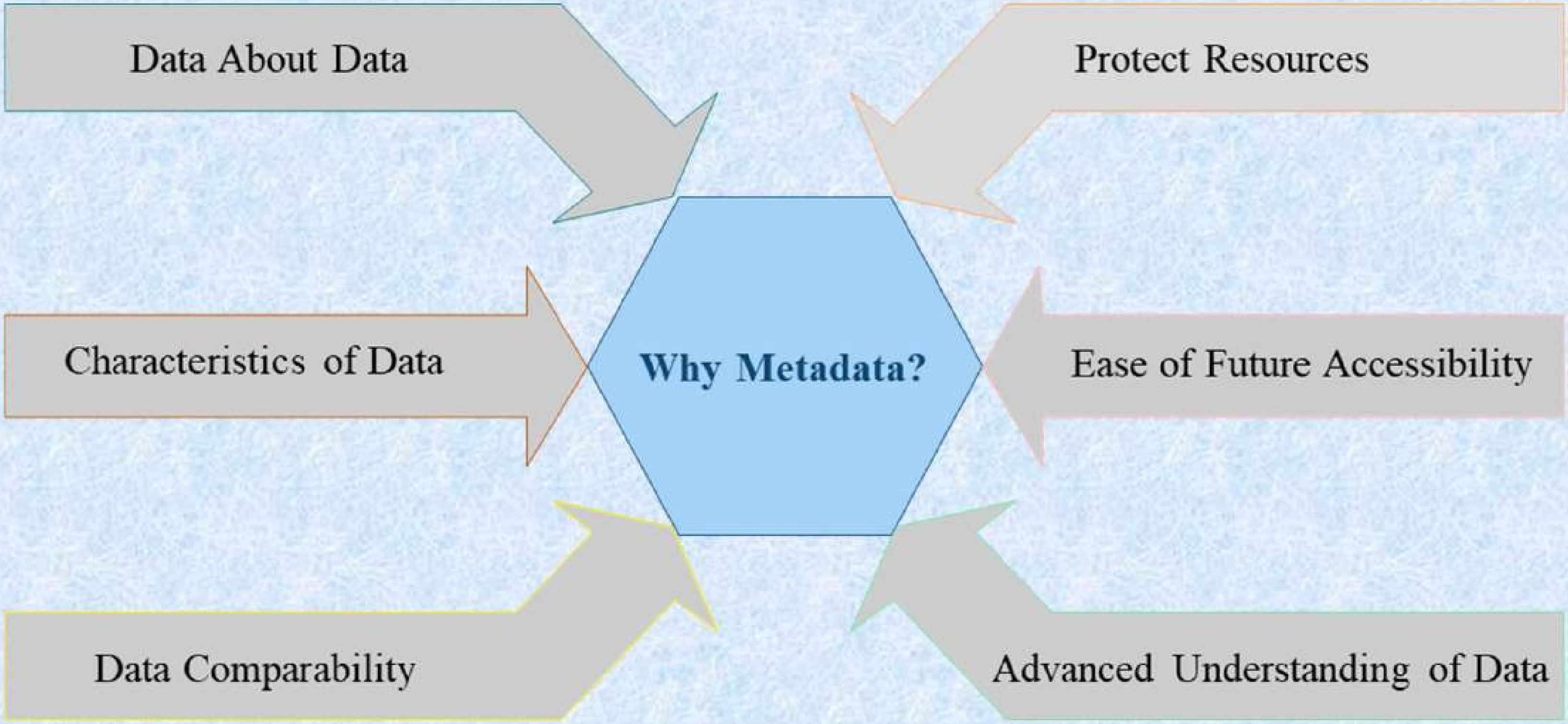
- Ensure that user reports contain all the information required by users to assess whether statistical outputs are fit for the purposes they have in mind.

What is metadata?

- It is structured information or documentation about data which:
  - informs users about the content, quality and condition of data;
  - describes the structure of datasets, explains, locates, or makes it easier to retrieve, use, or manage data;
  - provides information on the processes of data production;
  - guides on proper usage or interpretation of data.

- “Data that define and describe other data” (ISO definition)
- “Information about information” (Dion, 2006)

Why Metadata



Why Metadata?

Data About Data

Characteristics of Data

Data Comparability

Protect Resources

Ease of Future Accessibility

Advanced Understanding of Data

What does metadata do?

- Assist in retrieving and processing data
- Support correct use of data
- Provide transparency in data
- Enhance interoperability
- Improve archiving, preservation, institutional memory over time

Types of metadata

- **Structural metadata**
  - provide information about the structure of the dataset
  - act as identifiers and descriptors of the data, making it possible to properly identify, retrieve and browse the data
- **Reference metadata**
  - allow a thorough understanding and interpretation of the corresponding statistical data
  - Describe the concepts, definitions, methodology and quality of data; production and dissemination process, data access conditions, release policy, confidentiality, etc

Metadata standards (cont'd)

Two international metadata standards are becoming well established:

- SDMX (Statistical Data and Metadata Exchange)

  a number of international agencies have endorsed SDMX; supported by the UN Statistical Commission

-DDI (Data Dissemination Initiative)

  Microdata Management Toolkit of WB uses the DDI metadata standard

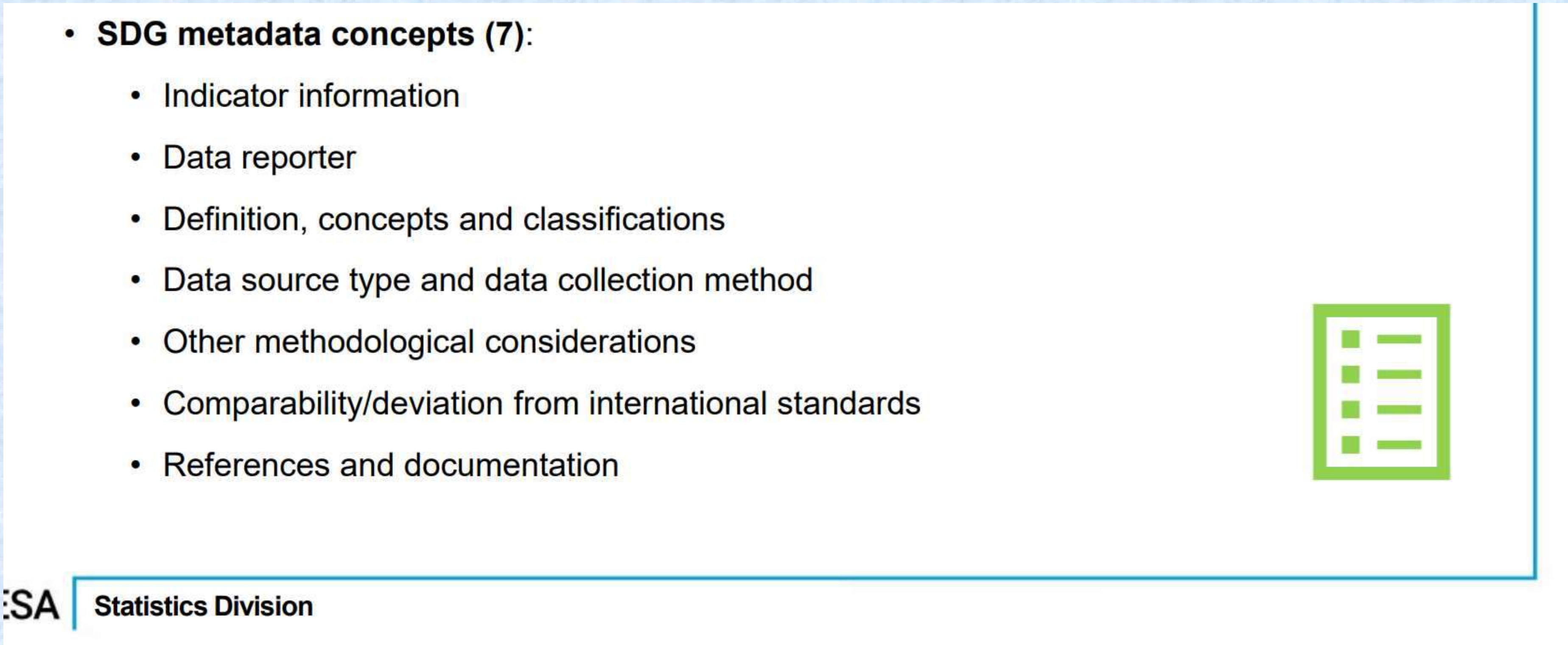
Users of metadata

Metadata support the knowledge of potential user of statistical information. The major users include:

- Users of statistical data
- Producers of statistical data
- Researchers on the development of statistical systems

Metadata Structure: International Organisations

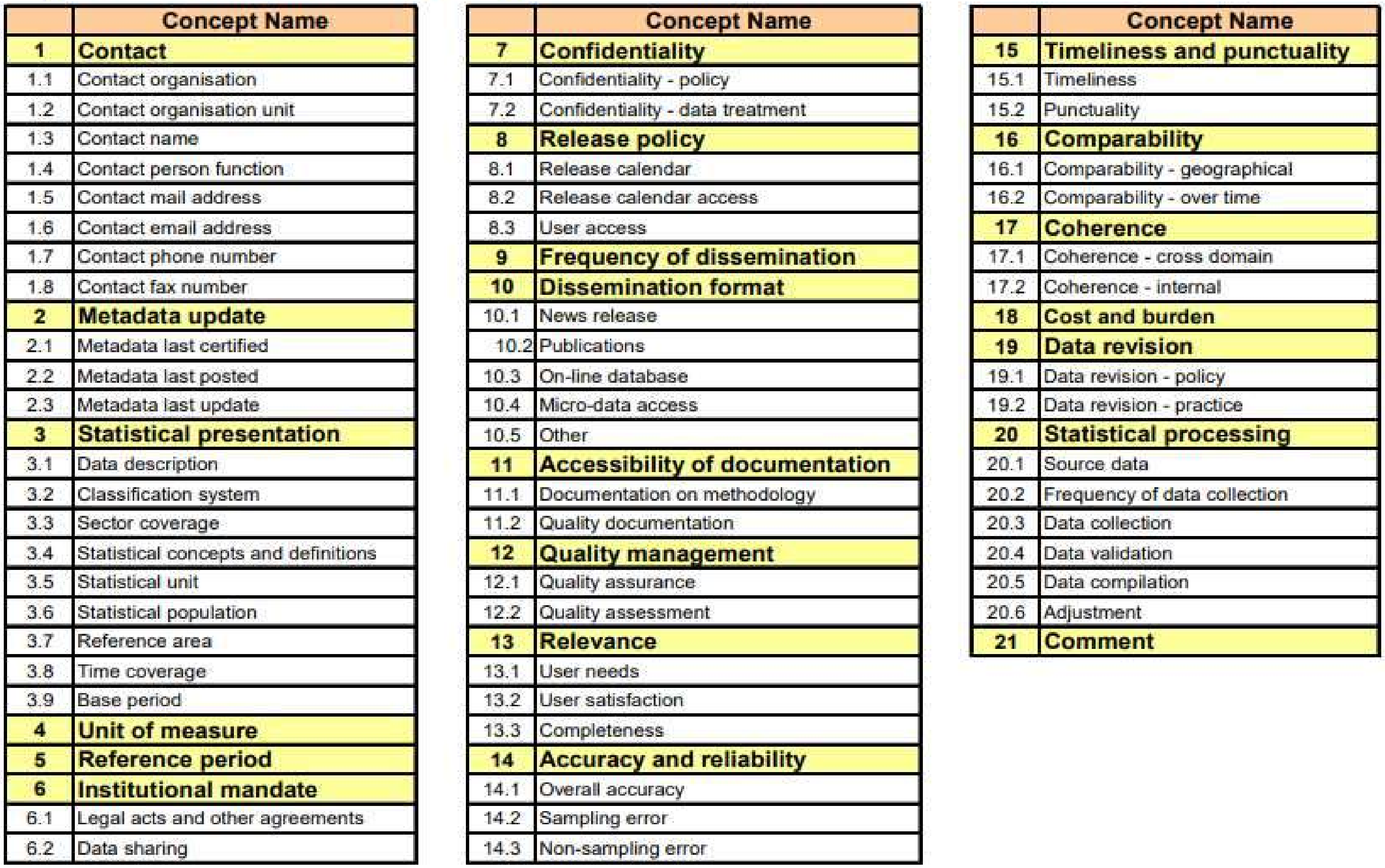
UN SDG Metadata Concept



- **SDG metadata concepts (7)**:
  - Indicator information
  - Data reporter
  - Definition, concepts and classifications
  - Data source type and data collection method
  - Other methodological considerations
  - Comparability/deviation from international standards
  - References and documentation

ESA | **Statistics Division**

Figure 2. The Euro SDMX Metadata Structure



	Concept Name
1	**Contact**
1.1	Contact organisation
1.2	Contact organisation unit
1.3	Contact name
1.4	Contact person function
1.5	Contact mail address
1.6	Contact email address
1.7	Contact phone number
1.8	Contact fax number
2	**Metadata update**
2.1	Metadata last certified
2.2	Metadata last posted
2.3	Metadata last update
3	**Statistical presentation**
3.1	Data description
3.2	Classification system
3.3	Sector coverage
3.4	Statistical concepts and definitions
3.5	Statistical unit
3.6	Statistical population
3.7	Reference area
3.8	Time coverage
3.9	Base period
4	**Unit of measure**
5	**Reference period**
6	**Institutional mandate**
6.1	Legal acts and other agreements
6.2	Data sharing

	Concept Name
7	**Confidentiality**
7.1	Confidentiality - policy
7.2	Confidentiality - data treatment
8	**Release policy**
8.1	Release calendar
8.2	Release calendar access
8.3	User access
9	**Frequency of dissemination**
10	**Dissemination format**
10.1	News release
10.2	Publications
10.3	On-line database
10.4	Micro-data access
10.5	Other
11	**Accessibility of documentation**
11.1	Documentation on methodology
11.2	Quality documentation
12	**Quality management**
12.1	Quality assurance
12.2	Quality assessment
13	**Relevance**
13.1	User needs
13.2	User satisfaction
13.3	Completeness
14	**Accuracy and reliability**
14.1	Overall accuracy
14.2	Sampling error
14.3	Non-sampling error

	Concept Name
15	**Timeliness and punctuality**
15.1	Timeliness
15.2	Punctuality
16	**Comparability**
16.1	Comparability - geographical
16.2	Comparability - over time
17	**Coherence**
17.1	Coherence - cross domain
17.2	Coherence - internal
18	**Cost and burden**
19	**Data revision**
19.1	Data revision - policy
19.2	Data revision - practice
20	**Statistical processing**
20.1	Source data
20.2	Frequency of data collection
20.3	Data collection
20.4	Data validation
20.5	Data compilation
20.6	Adjustment
21	**Comment**

MoSPI’s Metadata Structure

Brief Concepts of Metadata Structure

Item No	Concept name	Item No	Concept name	Item No	Concept name
1	**Contact**	3.2	Data sharing	7.2	Coherence
1.1	Contact Organisation	3.3	Release policy	**8**	**Statistical processing**
1.2	Compiling Agency	3.4	Release calendar	9.1	Source data type
1.3	Contact Details	3.5	Frequency of dissemination	9.2	Frequency of data collection
2	**Statistical presentation and description**	3.6	Data access	9.3	Data collection method
2.1	Data description	**4**	**Quality management**	9.4	Data validation
2.2	Classification system	4.1	Documentation on methodology	9.5	Data compilation
2.3	Sector coverage	4.2	Quality documentation	**9**	**Metadata update**
2.4	Statistical concepts and definitions	4.3	Quality assurance	9.1	Metadata last posted
2.5	Statistical unit	4.4	Quality assessment	9.2	Metadata last update
2.6	Statistical population	**5**	**Accuracy and reliability**		
2.7	Reference Period	5.1	Sampling error		
2.8	Base period	**6**	**Timeliness**		
2.9	Data Confidentiality	6.1	Timeliness		
3	**Institutional mandate**	**7**	**Coherence and comparability**		
3.1	Legal acts and other agreements	7.1	Comparability – over time		

Elements of Metadata Structure: Comparison amongst selected NSOs

Detailed Concepts of Metadata

Contact

Back to Brief Concepts

Item No	Concept name	Definition	Guidelines
1	**Contact**	**Individual or organisational contact points for the data or metadata, including information on how to reach the contact points.**	
1.1	Contact Organisation	The name of the organisation for data or metadata.	Provide the full name (not just acronym) of the organisation responsible for the processes and outputs (data and metadata) that are the subject of the report
1.2	Compiling Agency	Organisation collecting and/or elaborating the data being reported	Provide the full name of the Division under the organisation responsible for the processes and outputs (data and metadata) that are the subject of the report
1.3	Contact Details	The details of the contact points for the data or metadata.	Provide contact details of contact point(s) in following format: a. Organisation Name owning the processes and outputs: b. Author (if different from (a)) c. Disseminating Agency (if different from (a) and (b)) d. Name (first and last names): e. Designation: f. Postal address: g. email address (preferably designation based): h. Contact number: i. Fax number: If more than one name is provided, the details of main contact should be indicated. If the author of the report is different from the person(s) responsible for process and its outputs, provide this name also with his/her details

Statistical Presentation and Description

Back to Brief Concepts

2	**Statistical presentation and description**	**Description of the disseminated data which can be displayed to users as tables, graphs or maps**	
2.1	Data description	Main characteristics of the data set, referring to the data and indicators disseminated.	Describe briefly the main characteristics of the data in an easily and quickly understandable manner, referring to the main variables disseminated.
2.2	Classification system	Arrangement or division of objects into groups based on characteristics which the objects have in common	List all classifications and breakdowns that are used in the data (with their detailed names) and provide links (if publicly available). Type of dis-aggregation available in the data sets - for example rural-urban, male-female, etc. and whether data is available at the sub-national level or not, should be clearly specified.
2.3	Sector coverage	Main economic or other sectors	List the main economic or other sectors covered by the data and the size classes used, for example, Health/ Education/ Manufacturing etc
2.4	Statistical concepts and definitions	Statistical characteristics of statistical observations, variables	Define and describe briefly the main statistical variables that have been observed or derived. Indicate their types.
2.5	Statistical unit	Entity for which information is sought and for which statistics are ultimately compiled.	Define the type of statistical unit about which data are collected, e.g. enterprise, household, etc.
2.6	Statistical population	The total population of a defined class of people, objects or events	Define the target population of statistical units for which information is sought
2.7	Reference Period	The length of time for which data are available	State the time period(s) for which data is collected
2.8	Base period	The period of time used as the base of an index number, or to which a constant series refers.	Note that this concept applies only to certain types of outputs, such as indexes, for which a base period is defined and used. Eg, CPI, IIP, etc.
2.9	Data Confidentiality	Rules applied for treating the datasets to ensure statistical confidentiality and prevent unauthorised disclosure.	Describe the procedures that are used in protecting confidentiality.

Back to Brief Concepts

Institutional Mandate

3	Institutional mandate	Law, set of rules or other formal set of instructions assigning responsibility as well as the authority to an organisation for the collection, processing, and dissemination of statistics	
3.1	Legal acts and other agreements	Legal acts or other formal or informal agreements that assign responsibility as well as the authority to an agency for the collection, processing, and dissemination of statistics	State the national legal acts and/or other reporting agreements
3.2	Data sharing	Arrangements or procedures for data sharing and coordination between data producing agencies.	Describe the arrangements, procedures or agreements to facilitate data sharing and exchange between data producing agencies within the national statistical system
3.3	Release policy	Rules for disseminating statistical data to all interested parties	State if the release of the products is governed by some policy etc.
3.4	Release calendar	The schedule of statistical release dates.	State whether there is a release calendar for the statistical outputs from the process being reported, and if so, whether this calendar is publicly accessible and if yes, give a link or reference.
3.5	Frequency of dissemination	The time interval at which the statistics are disseminated over a given time period.	State the frequency with which the data are disseminated, e.g. monthly, quarterly, yearly.
3.6	Data access	The conditions and modalities by which users can access, use and interpret data	State the conditions and link on website from where the user can access the data For easy access of users, following details should also be mentioned about the dataset: **Title:** Name by which the data is known **Dataset Edition:** Edition of data (ex: first, second, final etc) **Dataset Reference data type:** Type of data entered in the field (ex: .txt, .dbf, .xls) **Presentation Format:** Presentation format of the data (ex: document, map, table, etc.) **Dataset Language:** language of any text in the data **Status/Version:** How updated is the data?

Quality Management

Back to Brief Concepts

4.	**Quality management**	**Systems and frameworks in place within an organisation to manage the quality of statistical products and processes.**	
4.1	Documentation on methodology	Descriptive text and references to methodological documents available.	List reference metadata files, methodological papers, summary documents and handbooks relevant to the statistical process.
4.2	Quality documentation	Documentation on procedures applied for quality management and quality assessment.	List relevant quality related documents, for example, other quality reports, studies.
4.3	Quality assurance	All systematic activities implemented that can be demonstrated to provide confidence that the processes will fulfil the requirements for the statistical output.	Describe the procedures (such as use of a general quality management system based on ISO 9000 series) to promote general quality management principles in the organisation. Describe the quality assurance framework used to implement statistical quality principles. Describe the quality assurance procedures specifically applied to the statistical process for which the report is being prepared, for example training courses, process monitoring, benchmarking, assessments, and use of best practices. Describe any ongoing or planned improvements in quality assurance procedures.
4.4	Quality assessment	Overall assessment of data quality, based on standard quality criteria.	Summarise the results of the most recent quality assessments and cross reference to the chapters in the report where the results are presented in more detail.

Back to Brief Concepts

Accuracy and Reliability

5	Accuracy and reliability	Accuracy of data is the closeness of computations or estimates to the exact or true values that the statistics were intended to measure. Reliability of the data, defined as the closeness of the initial estimated value to the subsequent estimated value.	
5.1	Sampling error	That part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a subset of the population is enumerated.	If probability sampling is used: • for user reports-provide the range of variation of the A1 indicator among key variables at user report level of detail; • for producer reports-provide the range of variation of the A1 indicator among key variables at producer report level of detail; • indicate the impact of sampling error on the overall accuracy of the results; • state how the calculation of sampling error is affected by adjustments for nonresponse, misclassifications and other sources of uncertainty, such as outlier treatment. If non-probability sampling is used: provide an assessment of representativeness, a motivation for the invoked model for estimation and risk of sampling bias

Back to Brief Concepts

Timeliness and Punctuality

6	**Timeliness**	**The timeliness of the data collection release to be compiled.**	
6.1	Timeliness	Length of time between data availability, the event or phenomenon the data describe, and final release to its users.	Outline the reasons for the time lag, if any. Outline efforts to reduce time lag in future.

Coherence and Comparability

Back to Brief Concepts

7	Coherence and comparability	Adequacy of statistics to be reliably combined in different ways and for various uses and the extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics	
7.1	Comparability – over time	The extent to which statistics are comparable or reconcilable over time	Provide information on possible limitations in the use of data for comparisons over time. Distinguish three broad possibilities: 1. There have been no changes, in which case this should be reported. 2. There have been some changes but not enough to warrant the designation of a break in series. 3. There have been sufficient changes to warrant the designation of a break in series.
7.6	Coherence	The extent to which statistics are reconcilable with System of National Accounts.	For producer reports only. Where relevant, the results of comparisons with the System of National Account framework / Other Statistical Standards and feedback from System of National Accounts / Other Statistical Standards with respect to coherence and accuracy problems should be reported and should be a trigger for further investigation.

Statistical Processing

Back to Brief Concepts

8	Statistical processing	Any statistical processing undertaken to finalise the data	
8.1	Source data type	Characteristics and components of the raw statistical data used for compiling statistical aggregates.	Indicate if the data are based on a survey, administrative data, multiple data sources, or macro-aggregates. In the event of multisource or macro-aggregate processes describe each source. For each dataset from an administrative source, summarise the source, its primary purpose, and the most important data items acquired
8.2	Frequency of data collection	Time interval at which the source data are collected	Indicate the frequency of data collection (e.g. monthly, quarterly, annually, or continuous).
8.3	Data collection method	Method applied for gathering data for official statistics.	For each source of survey data: • describe the method(s) used to gather data from respondents; • annex or hyperlink the questionnaires(s). For each source of administrative data: • describe the acquisition process and how it was tested. For all sources: • describe the types of checks applied at the time of data entry.

Statistical Processing (Contd…)

Back to Brief Concepts

8.4	Data validation	Process of monitoring the results of data compilation and ensuring the quality of statistical results	Describe the procedures for checking and validating the source data and how the results are monitored and used. Describe the procedures for validating the aggregate output data (statistics) after compilation, including checking coverage and response rates, and comparing with data for previous cycles and with expectations. List other output datasets to which the data relate and outline the procedures for identifying inconsistencies between the output data and these other datasets
8.5	Data compilation	Operations performed on data to derive new information according to a given set of rules.	Describe the procedures for imputation, the most common reasons for imputation and imputation rates within each of the main strata. Describe the likely impact of imputation. Describe the procedures to derive new variables and to calculate aggregates and complex statistics. Describe the procedures for adjustment for non-response and the corrections to the design weights to account for differences in response rates. Describe the calculation of design weights, including calibration (if used). Describe the procedures for combining input data from different sources. Provide the ratio of the number of replaced values to the total number of values for a given variable. Specific reference to formula shall be made. The formula or mathematical equation used while computing different variables in the report may be described here in a structured format showing the Numerator; Denominator and Multiplier used for computing the same

Metadata Update

Back to Brief Concepts

9	Metadata update	The date on which the metadata element was inserted or modified in the database.	
9.1	Metadata last posted	Date of the latest dissemination of the metadata	The date when the complete set of metadata was last disseminated as a block should be provided (manually, or automatically by the metadata system).
9.2	Metadata last update	Date of last update of the content of the metadata.	The date when any metadata were last updated should be provided (manually, or automatically by the metadata system).

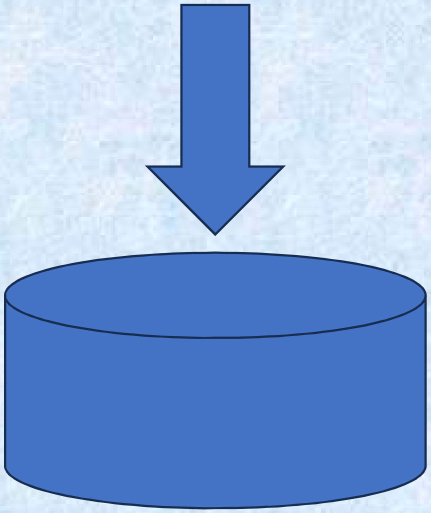
Metadata Structure for Statistical Products as Available in Public Domain

Use Cases of Metadata Compliance of MoSPI

What Next

- From ”paper” to database

- Information in the word document



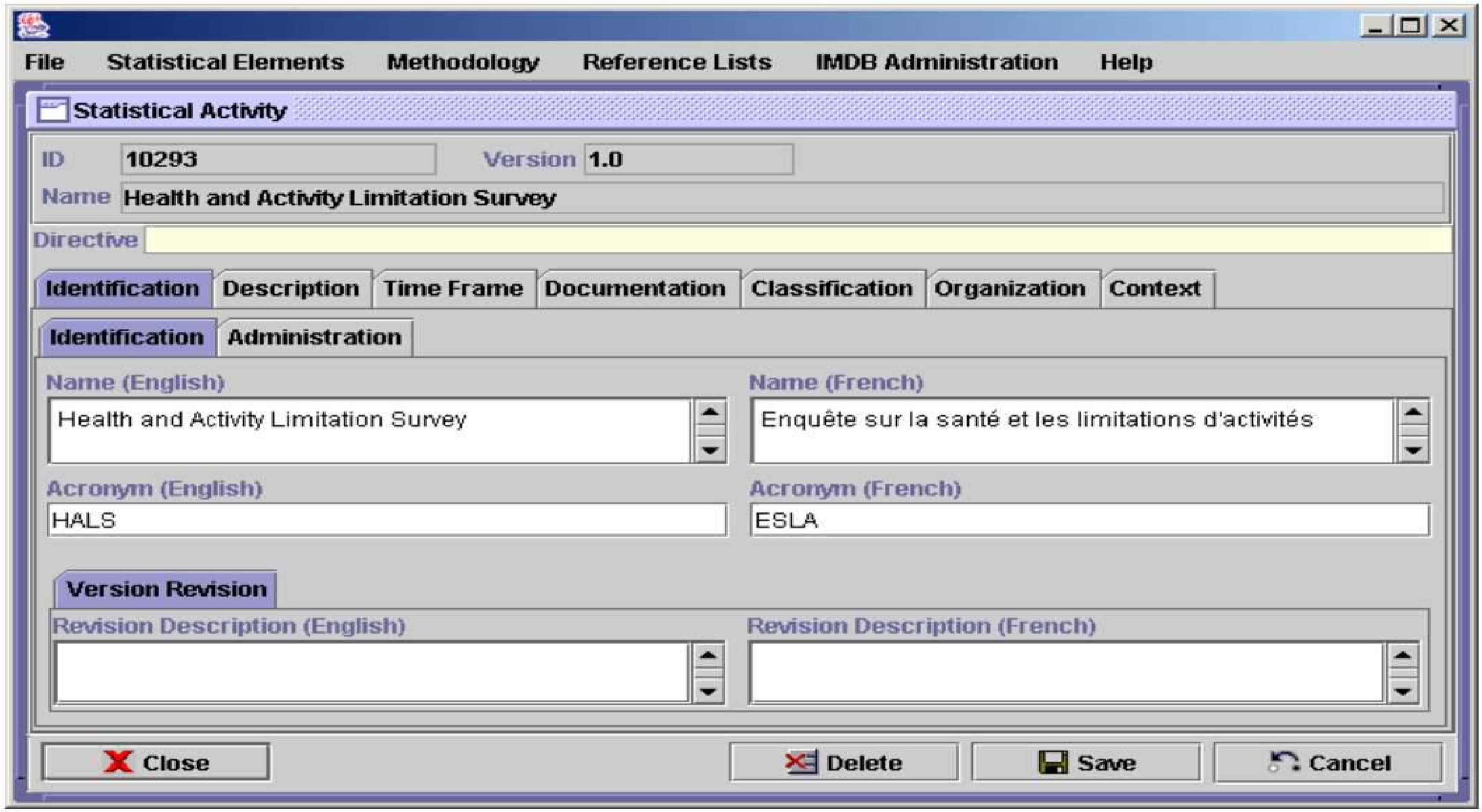
Revised NMDS

Introduction of Two new Concepts

- **2.3 International/National Standards Classification etc.**
- **8.6 Identifier of a statistical indicator**

NMDS

Loading metadata into India Integrated Metadata Base (IIMDB)


File
Statistical Elements
Methodology
Reference Lists
IMDB Administration
Help
Statistical Activity
Identification
Description
Time Frame
Documentation
Classification
Organization
Context
Identification
Administration
Version Revision
Close
Delete
Save
Cancel

ID: 10293

Version: 1.0

Name: Health and Activity Limitation Survey

Directive:

Name (English): Health and Activity Limitation Survey

Name (French): Enquête sur la santé et les limitations d'activités

Acronym (English): HALS

Acronym (French): ESLA

Revision Description (English):

Revision Description (French):

Thank You

National Data and Analytics Platform (NDAP)

13.08.2024

10100
11010
AI
0100
1010
11011

1. DATA ECOSYSTEM
2. ABOUT NDAP
3. KEY FEATURES OF NDAP
4. Other DATA Aggregators
5. State Data Analytics Platform
6. Way Forward

Statistic Graph
Statistic Graph
Statistic Graph
Statistic Graph
55%

01

Data Ecosystem

Data Ecosystem: Key Elements

Identification of Data Sources & Collection

Data Processing (transforming raw data into a usable format)

Data Dissemination

Data Analysis

Key challenges in the present ecosystem

Inaccessible

Public government data are difficult to find, understand, and download

Siloed

Datasets do not speak to each other across space, time and sector, which makes cross-sectoral analysis difficult

Cluttered

Existing platforms are complex, cluttered, unintuitive, and difficult to navigate

Static

Visualizations and analytics are largely static and inflexible; dashboards are not interactive

Lossy

Public versions of government data are often less-useful aggregates

NDAP addresses these challenges in the data ecosystem

Substantial amount of data available in the system without necessary discovery or means to draw insights

NDAP provides access to standardized data at a single platform; allowing users to search, merge, visualize & download datasets easily

What is NDAP?

A platform that has intuitive features given its user-centric approach

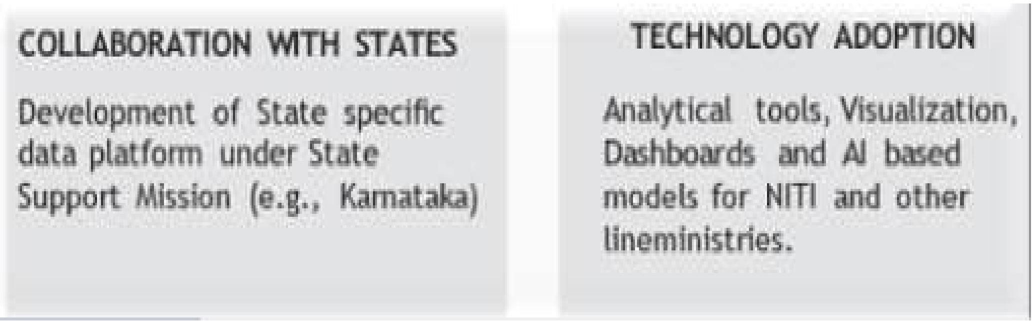
Streamlines published datasets with the mission to host the entire universe of identifiable published datasets (est. 5000) in a standardized schema at one platform making data inter-operable

What it covers

- Published government data

What it does not cover

- Individual Data
- Transactional Data
- Unpublished data

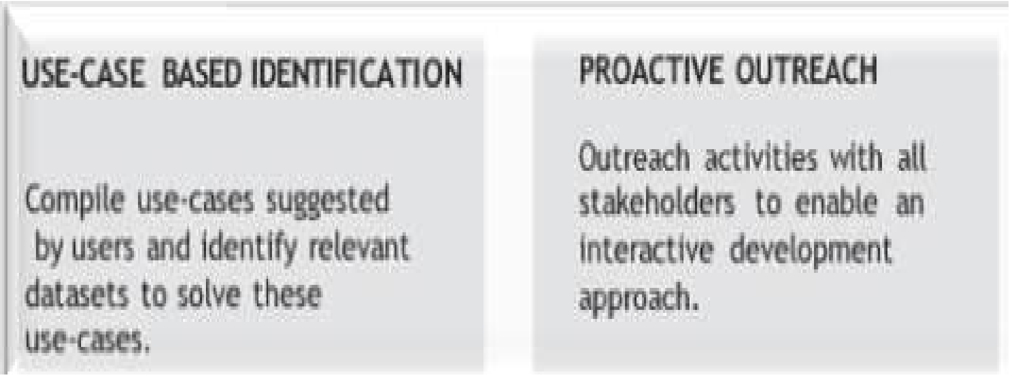


COLLABORATION WITH STATES

Development of State specific data platform under State Support Mission (e.g., Karnataka)

TECHNOLOGY ADOPTION

Analytical tools, Visualization, Dashboards and AI based models for NITI and other lineministries.



USE-CASE BASED IDENTIFICATION

Compile use-cases suggested by users and identify relevant datasets to solve these use-cases.

PROACTIVE OUTREACH

Outreach activities with all stakeholders to enable an interactive development approach.

Key features of NDAP

Standardized data

NDAP hosts published government data in machine-readable formats which are standardized using the LGD directory

Interoperability of data

Datasets on NDAP are interoperable, which can be merged on the platform using **in-built Merge tool** to do cross sectoral analysis

Feedback/Suggestion

Inbuilt process of time-bound redressal of user feedback

User-Friendly UI/UX and Intuitive Search engine

NDAP has simple, intuitive search and coherent organization of datasets to allow easy browsing

Interactive Visualization

Users can create maps, bar charts, line charts, pie charts, and scatter plots using any dataset and indicator on NDAP

Use Case based approach

168 use cases are hosted on the platform to guide users on data analytics

Current Status

168 Use Cases

Total Datasets on NDAP as on date

3735

53 Ministries

31 Sectors

3400+ Datasets

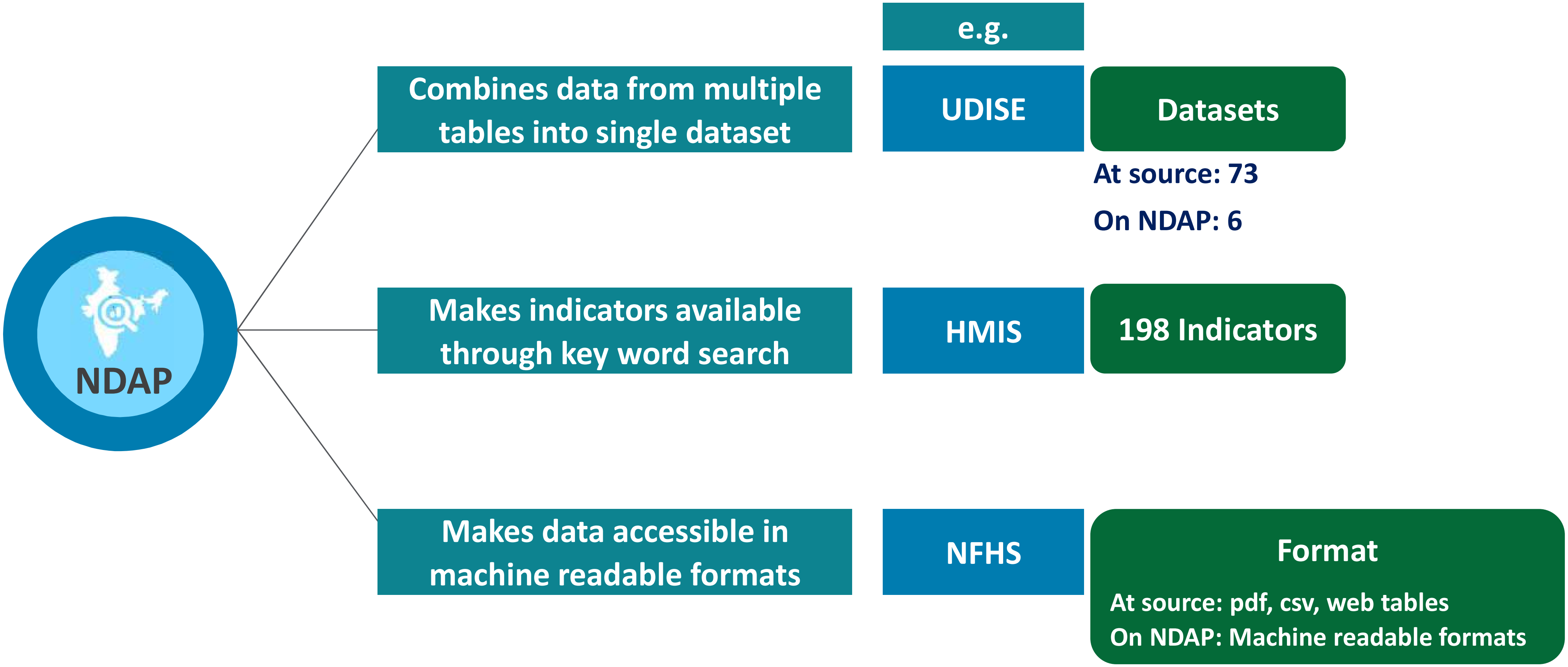
Tracking Users, Sessions, Queries …

Self-Service Visualizations

Vector DB with NLU Search

- Every quarter datasets are onboarded on NDAP as per agreement and priorities and feedback
- 4217 – No. of Datasets will be on NDAP by end of September 2024
- Data on NDAP is available through APIs

NDAP streamlines datasets



e.g.

Combines data from multiple tables into single dataset

UDISE

Datasets

At source: 73

On NDAP: 6

Makes indicators available through key word search

HMIS

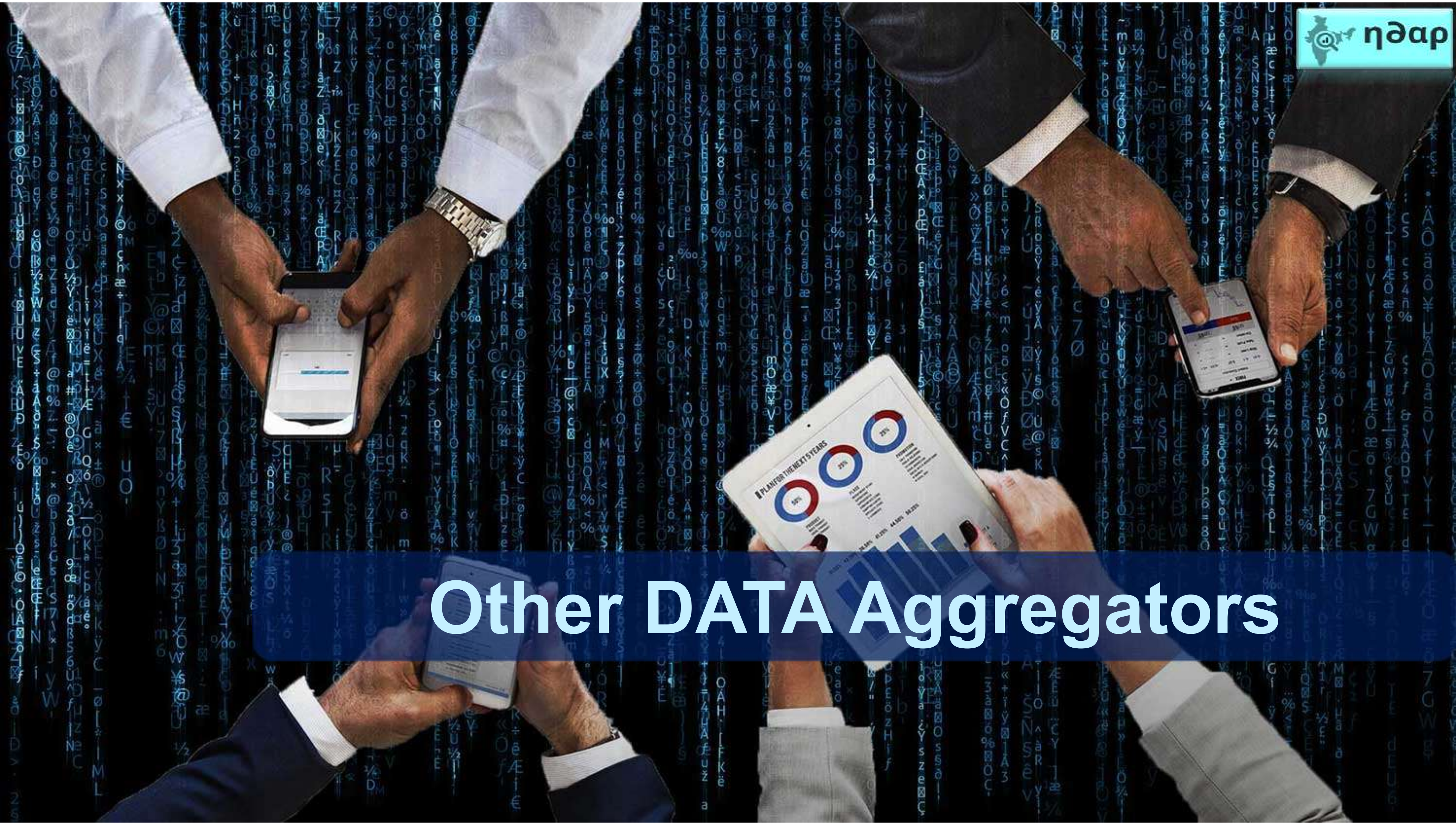
198 Indicators

Makes data accessible in machine readable formats

NFHS

Format

At source: pdf, csv, web tables
On NDAP: Machine readable formats


PLAN FOR THE NEXT 5 YEARS

Other DATA Aggregators

Major sources of data at national level

Platform	Owner	Data Coverage	Accessibility
Data.gov.in	NIC & MeitY	Central ministries and state specific data	Open access
Esankhyiki. mospi	MoSPI	Mostly statistical data	Open access
RBI	RBI	Indian Economy	Open as per Special Data Dissemination Standards (SDDS)
Ministries/ Departments	Central or State Government	Ministry or Department specific data	Open Access/ Login based
NDAP	**NITI Aayog**	**Published Government data** **Currently 3000+ datasets from 31 sectors**	**Open access**

NDAP - International Benchmarking (1/2)

FEATURES ⬇	**PLATFORMS ➡**	**NDAP**	**DATA USA**	**FRED**	**WORLD BANK DATA PORTAL**
Allows to Download datasets (including through APIs)		✓	✓	✓	✓
Mobile First approach		X	✓	✓	X
Release Calendar*		X	✓	✓	✓
Dataset Request		✓	✓	X	✓
GIS mapping		X	X	X	✓
News and Community updates		X	✓	✓	✓

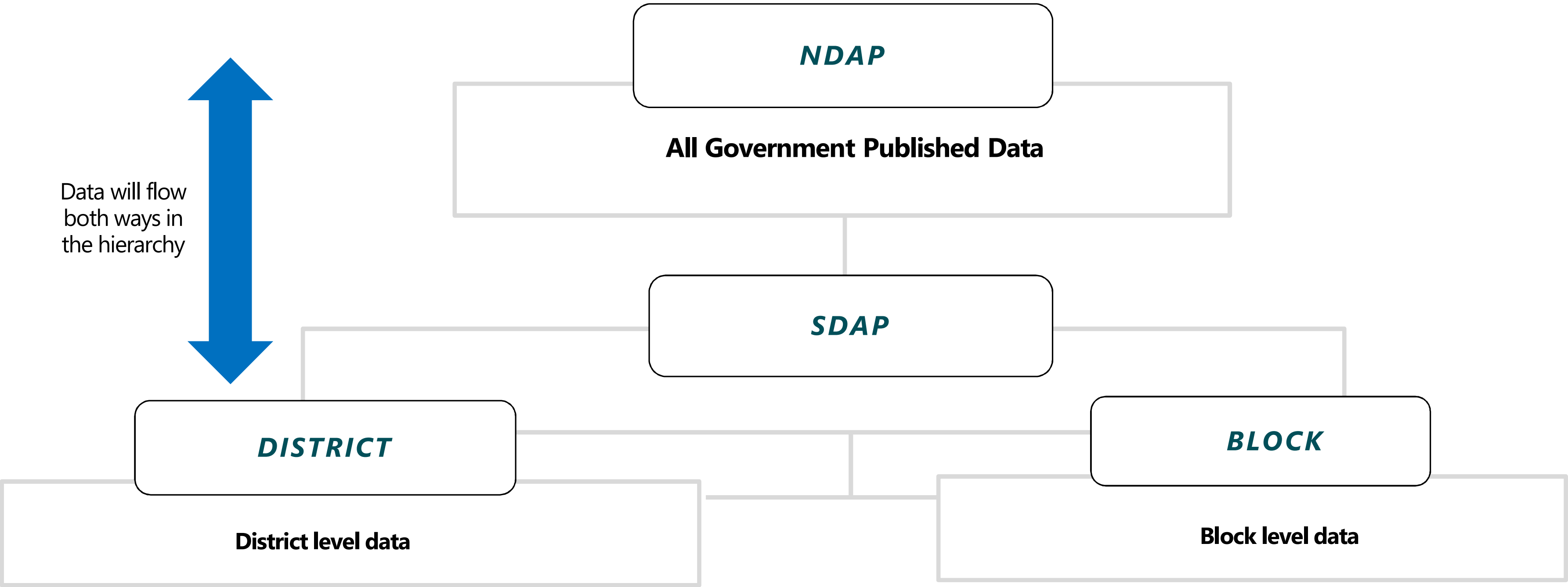
** NDAP is not the source of the data and it refreshes data in real-time where API is in place and within 20 days after updation at the source in other cases. Therefore, an advance release calendar cannot be given.*

NDAP - International Benchmarking (2/2)

FEATURES ⇩ PLATFORMS ⇨	NDAP	DATA USA	FRED	WORLD BANK DATA PORTAL
Search option	✓	✓	✓	✓
Merging Datasets	✓	✗	✓	✗
Visualization Options	✓	✓	✓	✓
Dataset Information (including detailed Metadata)	✓	✓	✓	✓
User Feedback and Dataset Suggestion	✓	✗	✓	✓
Share reports on social media	✓	✓	✓	✓
Real time data update*	✗	✗	✓	✓

** NDAP is not the source of the data and it refreshes data in real-time where API is in place and within 20 days after updation at the source in other cases.*

State Data Analytics Platform (SDAP)



Data will flow both ways in the hierarchy

NDAP

All Government Published Data

SDAP

DISTRICT

District level data

BLOCK

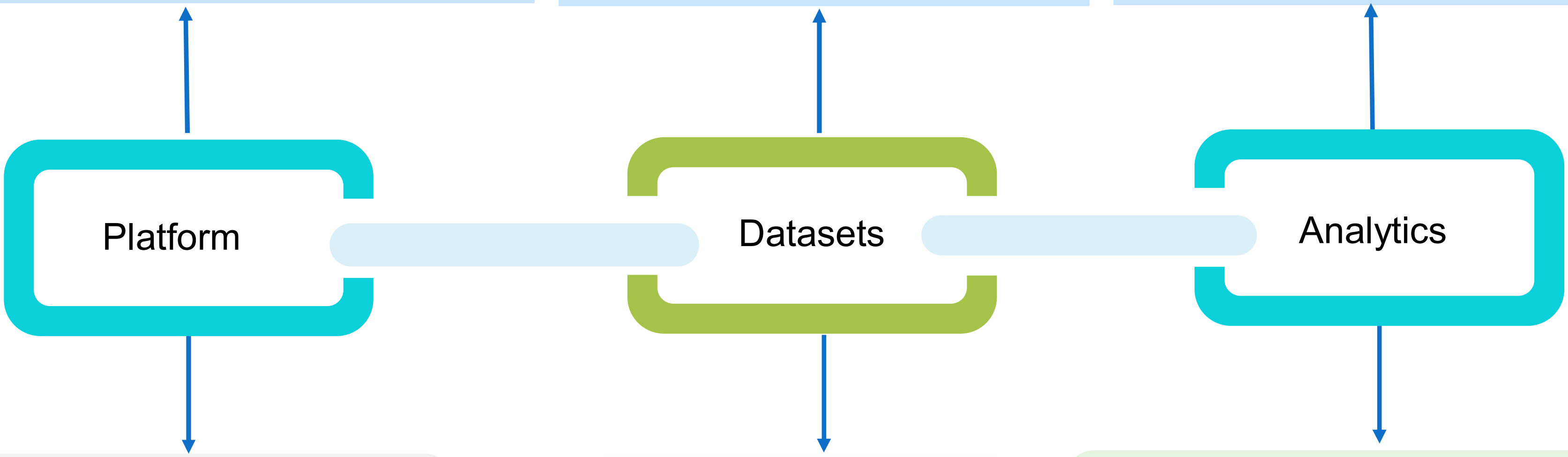
Block level data

Benefits to State

- **Customized State Data and Analytics Portal will be developed by NITI for respective state governments with all functionalities of NDAP**
- **User will be able to view both State and National level data on the same platform for doing cross-sectorial analysis**

- **Standardized, interoperable datasets** for policy planning, monitoring of Central/ State Government flagship schemes, prepare reports, visualization and dashboards.
- A State-specific **interface** having datasets with **all functionalities of NDAP**

- **Comparing data at pan-India level along with ease of access to all stakeholders/ users within and beyond the State boundary.**
- **Identify and solve use-cases suggested by various users through relevant datasets**



- Availability of State/ District/ Block/ Village level datasets on NDAP

- NDAP is enriched with more datasets from various sectors/ schemes

- Leverage the State specific datasets for creation of various **indices and dashboards** at NITI
- Opportunity to promote data analysis & **policy insights** at all levels of government

Benefits to NITI

Major Achievements

More than 3700 datasets from 53 Ministries across various 31 sectors have been onboarded on NDAP

SDAP (State Data and Analytics Platform) for 2 States, Karnataka and Meghalaya went live in Feb, 2023 and April, 2024 respectively

Workshop conducted with States in Lucknow, on 17th November 2023 on ‘Leveraging State Data Ecosystems for State and District Level Policy and Planning

PLAN FOR THE NEXT 5 YEARS

Way Forward

Initiatives under NDAP 2.0

Natural Language (NL) model based Analytical Layer

- When the user asks a question on the platform, the system will respond with appropriate output or response based on the data available on NDAP

Indicator Pages on NDAP

- Indicator pages will provide quick insights on the indicator and enable users to access commonly sought indicators and pre-made visualization and charts.
- Update UI/ UX of NDAP

Harmonized Microdata on NDAP

- Micro Data published by various government sources, harmonized at a single platform

Building State Data Analytics Platform (SDAP) for States

- Customized SDAP with all functionalities of NDAP along with unique state-specific requirements E.g. Karnataka Data Analytics Platform (KADAP) & Meghalaya Data Analytics Platform (MegDAP)

Thank You

10100
11010
AI
0100
1010
11011

Ministry of Electronics and Information Technology
Government of India

Digital India
Power To Empower

Open Government Data (OGD) Ecosystem in India

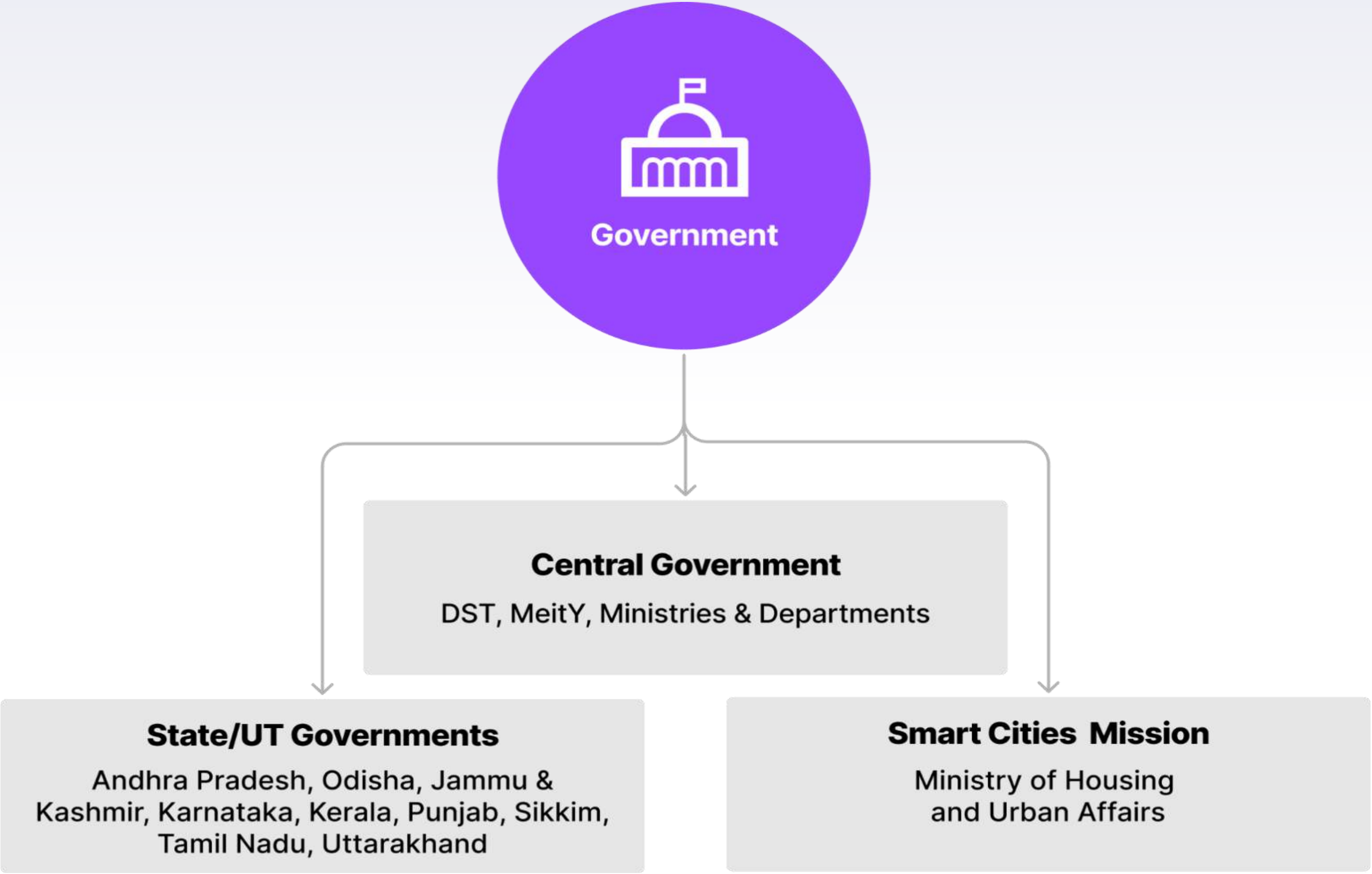
https://data.gov.in

Alka Misra
Deputy Director General
National Informatics Centre
Ministry of Electronics and Information Technology

National Informatics Centre

data.gov.in
Open Government Data (OGD) Platform Indi

OGD Ecosystem



Government

Central Government

DST, MeitY, Ministries & Departments

State/UT Governments

Andhra Pradesh, Odisha, Jammu & Kashmir, Karnataka, Kerala, Punjab, Sikkim, Tamil Nadu, Uttarakhand

Smart Cities Mission

Ministry of Housing and Urban Affairs

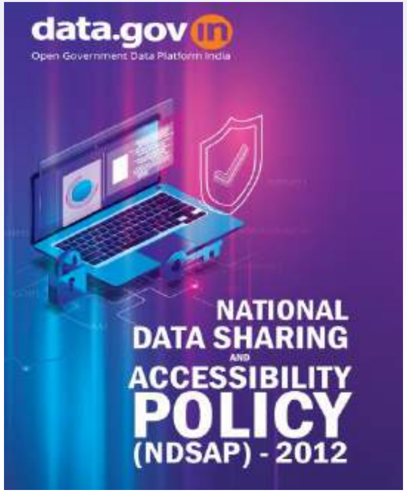
Launch of OGPL with Indo-US Partnership

Open Government Platform (OGPL)
Government
Citizens
Data Management
Data Catalogs
Workflow
Standard Metadata
Validation
Analytics
Data Portal
पत्र सूचना कार्यालय
PRESS INFORMATION BUREAU
भारत सरकार
GOVERNMENT OF INDIA
ROBERT BLAKE

Brainstorming sessions to enhance Data Ecosystem

SCOPE
data.gov.in
ational Conference on
data.gov
Chief Guest
i Shankar
data.gov.in
Workshop on
Data Driven Decision Making
for Chief Data Officers
ormation Technology (DeitY)
nformation Technology
NIC
DR. RAJENDRA KUMAR
MS. NEETA VERMA
OPEN GOVERNMENT DATA
& CIVIC ENGAGEMENT
data.gov.in
By Shri Kapil Si
Hon'ble Minister of
Communications & Information T
and Minister of Law & Jus
18th FEBR Y 2014
SCOPE CONVE N CENTR
Information Technology (DeitY)
& Info n Technology

National Data Sharing and Accessibility Policy (NDSAP)



data.gov.in
Open Government Data Platform India

NATIONAL DATA SHARING AND ACCESSIBILITY POLICY (NDSAP) - 2012

Formulated by D/o Science and Technology | Implemented by NIC, MeitY

- NDSAP Implementation Guidelines
- Development of the Platform – https://data.gov.in

Data Published by

- Ministries/Departments/Organizations facilitate proactive Release of Data
- For all the data and information created, generated, collected and archived using public funds provided by Government of India

GOVERNMENT OPEN DATA LICENCE - INDIA
REDEFINING OPENNESS

Government Open Data License (GODL)

- GODL – India developed to give a legal frame work to data consumers

  "All users are provided a worldwide, royalty free, non-exclusive license to use, adapt, publish, create derivative works (including products and services) for all lawful commercial and non-commercial purposes."

Journey of OGD Platform: 2012-2014

2012

- **Policy: 1.** Gazette Notification of NDSAP **2.** Implementation Guidelines Drafted through Consultations with Key Stakeholders
- **Software:** Alpha Release of OGD Platform

2013

- **Software: 1.** Beta Release of OGD Platform **2.** Visualization Engine Launch
- **Events: 1.** 12th Plan Hackathon **2.** In Pursuit of an Idea **3.** Open Data Apps Challenge **4.** CMA Hackathon

2014

- **Policy:** Implementation Guidelines Updated
- **Software: 1.** Stable Version of OGD Platform **2.** Data Access via APIs Started **3.** Sharing of Community-created Apps Started
- **Events: 1.** National Conference of Open Data and Open API **2.** Code for Honor

Journey of OGD Platform: 2015-2018

2015

- **Policy:** Implementation Guidelines Updated
- **Software:** Community Portal Launch
- **Events:** Workshops with Data Contributors & Chief Data Officers

2016

- **Policy: 1.** Government Open Data License – India Approved **2.** Implementation Guidelines Updated
- **Software: 1.** Events Portal Launch, **2.** SaaS Instance Deployments for Sikkim & Surat
- **Events:** Workshops with Chief Data Officers

2017-2018

- **Policy:** Gazette Notification of Government Open Data License – India
- **Software:** SaaS Instance Deployment for Tamil Nadu
- **Events: 1.** #OpenGovDataHack **2.** National Award to Students & Startups

Journey of OGD Platform: 2019-2024

2019

Software: SaaS Instance Deployment for Smart Cities Mission, Odisha, Karnataka & Kerala

Events: 1. 2nd Edition of #OpenGovDataHack (Chandigarh, Raipur, Rajkot, Kozhikode, etc.)
2. DARPG Hackathon for AI-driven Public Grievance Redressal Solution

2020-2021

Software: 1. Transition to Microservices Architecture & Data Migration **2.** Catalog-level APIs Feature Added **3.** SaaS Microservices) Instance Deployment for Punjab

2022-2024

Software: 1. Bulk Download Feature **2.** Migration of Earlier SaaS Instances to Microservices Architecture **3.** SaaS (Microservices) Instance Deployments for AP, J&K & UK **4.** Large Datasets, such as AgMarkNet, eShram & Kisan Call Center (KCC), Published via APIs

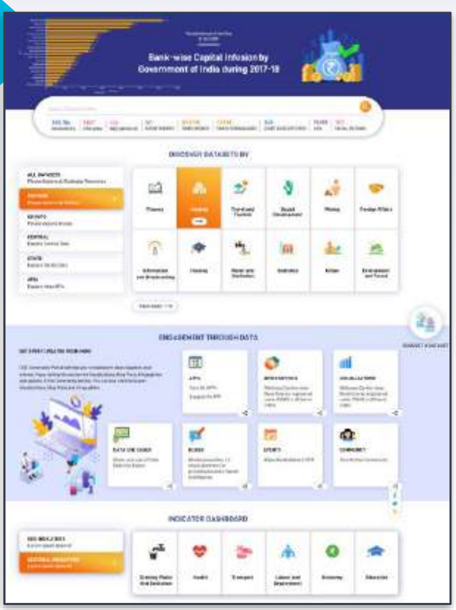
Events: 1. DARPG Hackathon on Data-driven Innovation for Citizen Grievance Redressal **2.** GSTN Analytics Hackathon on Developing a Predictive Model (Upcoming)

data.gov.in
Open Government Data (OGD) Platform India

OGD Platform Evolution

Digital India
Power To Empower

2021 (OGD 2.0)



Bank-wise Capital Infusion by Government of India during 2017-18

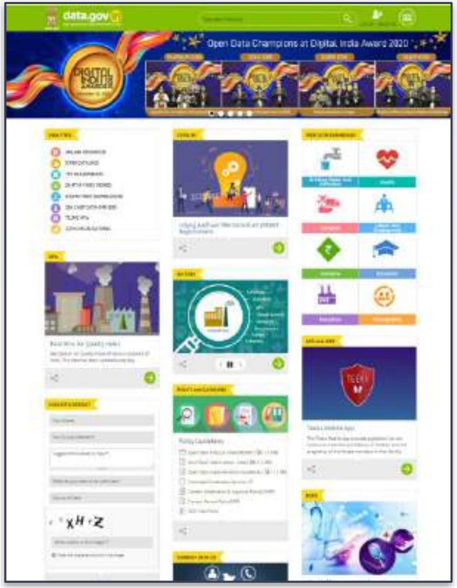
DISCOVER DATASETS BY

ENGAGEMENT THROUGH DATA

INDICATOR DASHBOARD

OGD 2.0 - Micro Services Based Architecture Leveraging Cloud Technology (MeitY Project)

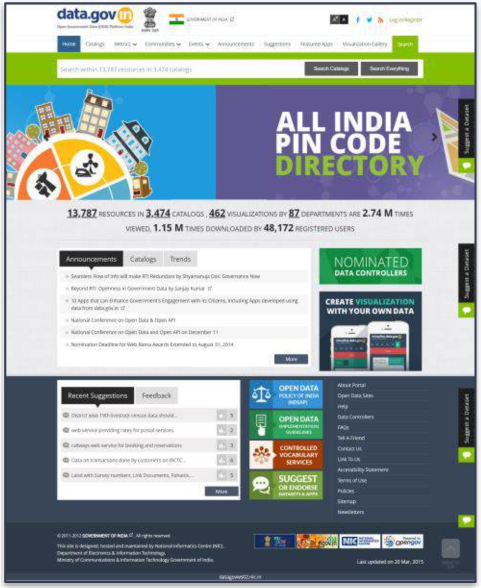
2017 - 2020



data.gov in

Open Data Champions at Digital India Award 2020

2015



data.gov in

Home | Catalogs | Metrics | Communities | Events | Announcements | Suggestions | Featured Apps | Visualization Gallery | Search

ALL INDIA PIN CODE DIRECTORY

13,787 RESOURCES IN 3,474 CATALOGS , 462 VISUALIZATIONS BY 87 DEPARTMENTS ARE 2.74 M TIMES VIEWED, 1.15 M TIMES DOWNLOADED BY 48,172 REGISTERED USERS

Announcements | Catalogs | Trends

- Seamless Flow of Info will make RTI Redundant by Shyamanuja Das: Governance Now
- Beyond RTI: Openness in Government Data by Sanjay Kumar
- 10 Apps that can Enhance Government's Engagement with its Citizens, including Apps developed using data from data.gov.in
- National Conference on Open Data & Open API
- National Conference on Open Data and Open API on December 11
- Nomination Deadline for Web Ratna Awards Extended to August 31, 2014

NOMINATED DATA CONTROLLERS

CREATE VISUALIZATION WITH YOUR OWN DATA

Recent Suggestions | Feedback

OPEN DATA POLICY OF INDIA (NDSAP)

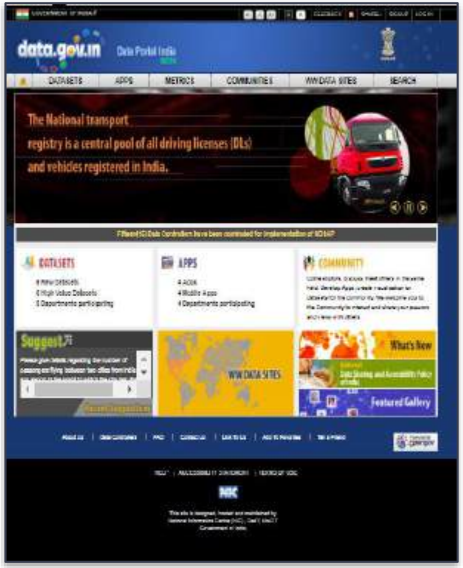
OPEN DATA IMPLEMENTATION GUIDELINES

CONTROLLED VOCABULARY SERVICES

SUGGEST OR ENDORSE DATASETS & APPS

Roll Out & Promotion of OGD Platform (MeitY Project 2014-2020)

2012



data.gov.in Data Portal India

DATASETS | APPS | METRICS | COMMUNITIES | WW DATA SITES | SEARCH

The National transport registry is a central pool of all driving licenses (DLs) and vehicles registered in India.

DATASETS

APPS

COMMUNITY

Suggest

WW DATA SITES

What's New

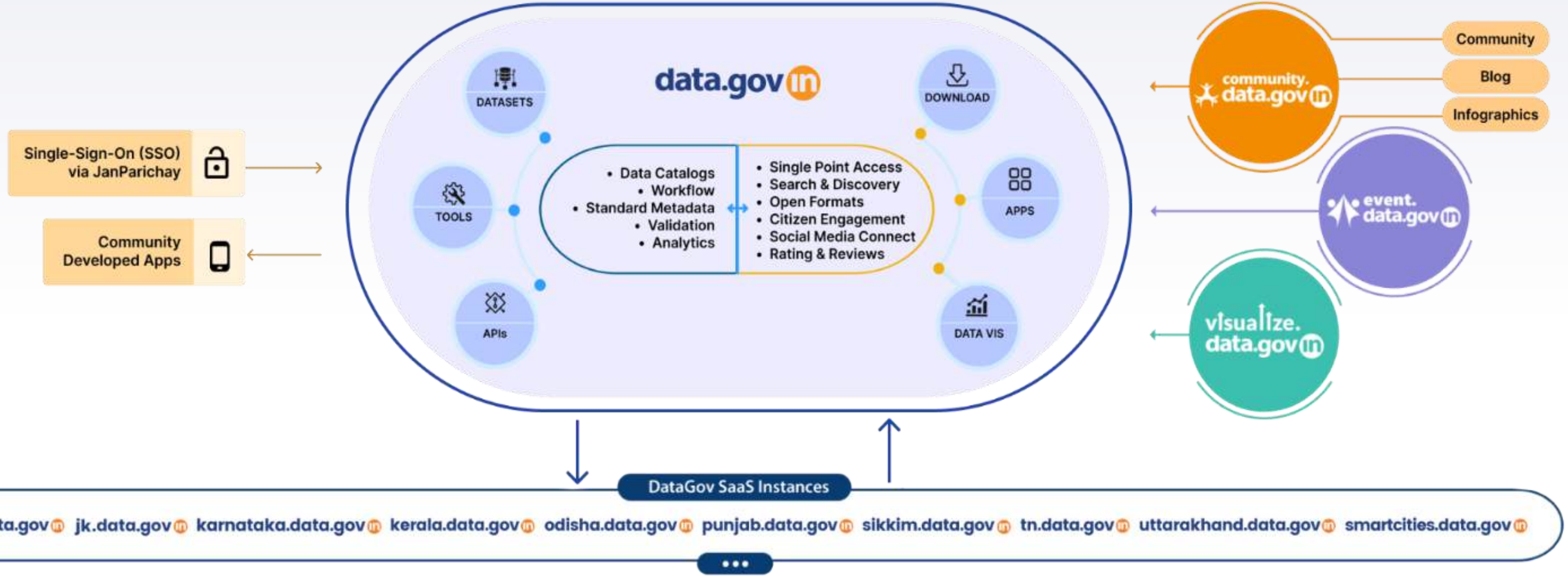
Featured Gallery

NIC Launched OGD Platform in 2012 using its own fund

data.gov in
Open Government Data (OGD) Platform India

NIC National Informatics Centre

OGD Platform Ecosystem (1/2)



Single-Sign-On (SSO) via JanParichay

Community Developed Apps

data.gov.in

DATASETS

TOOLS

APIs

DOWNLOAD

APPS

DATA VIS

- Data Catalogs
- Workflow
- Standard Metadata
- Validation
- Analytics

- Single Point Access
- Search & Discovery
- Open Formats
- Citizen Engagement
- Social Media Connect
- Rating & Reviews

community.data.gov.in

Community

Blog

Infographics

event.data.gov.in

visualize.data.gov.in

DataGov SaaS Instances

ap.data.gov.in jk.data.gov.in karnataka.data.gov.in kerala.data.gov.in odisha.data.gov.in punjab.data.gov.in sikkim.data.gov.in tn.data.gov.in uttarakhand.data.gov.in smartcities.data.gov.in

•••

502,174+	12,728 +	35.01+ M	240,060 +	10.28 + M	571 +	3,160 +
DATASETS	CATALOG	TIMES VIEWED	API	TIMES DOWNLOADED	CHIEF DATA OFFICERS	VISUALIZATIONS

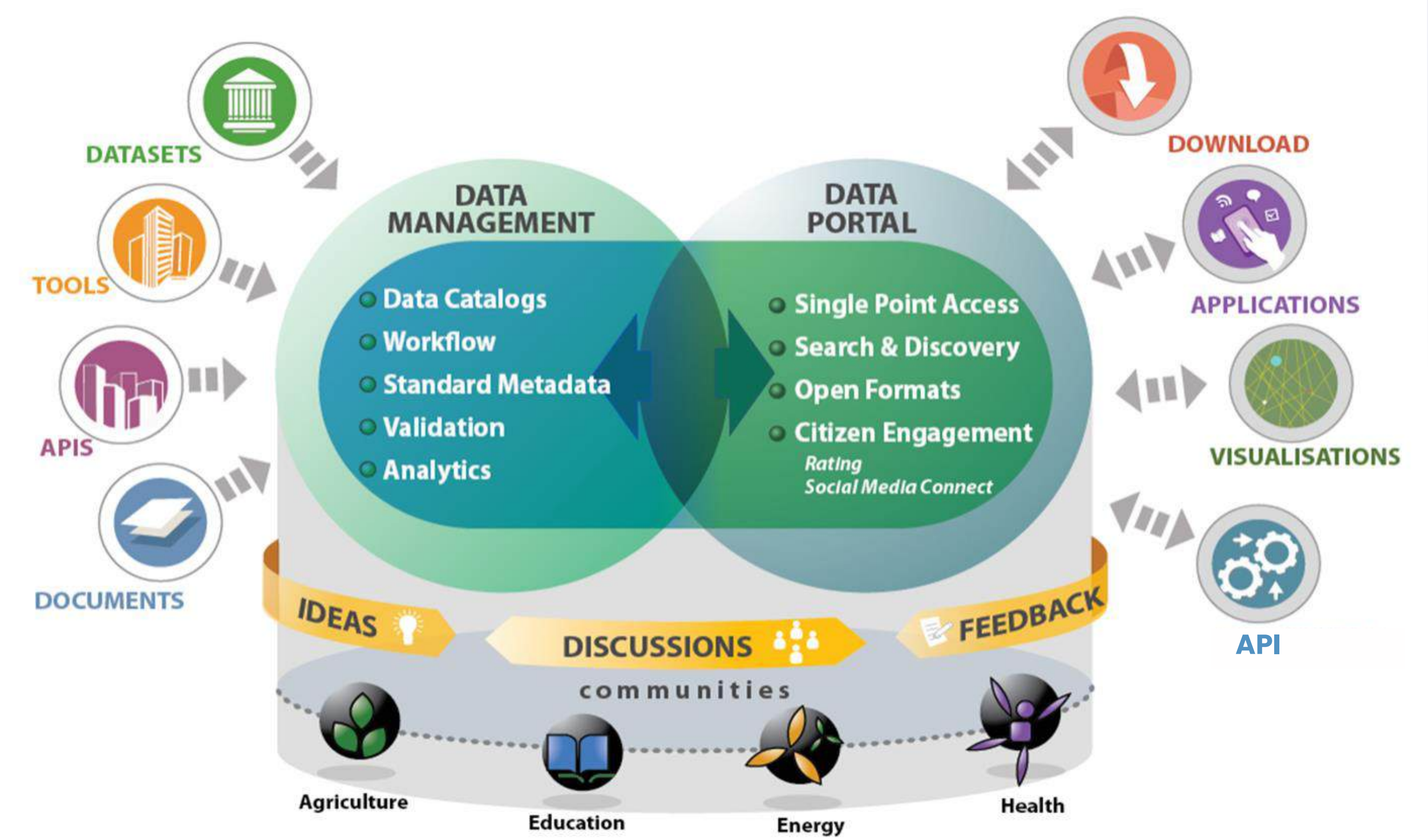
Registered Users
510,890 +

Data Portal India
19K + Followers

@DataPortalIndia
148.08K + followers

Last Updated on 31 July 2024

OGD Platform Ecosystem (2/2)



DATASETS

TOOLS

APIS

DOCUMENTS

DATA MANAGEMENT

- Data Catalogs
- Workflow
- Standard Metadata
- Validation
- Analytics

DATA PORTAL

- Single Point Access
- Search & Discovery
- Open Formats
- Citizen Engagement
  Rating
  Social Media Connect

DOWNLOAD

APPLICATIONS

VISUALISATIONS

API

IDEAS

DISCUSSIONS

FEEDBACK

communities

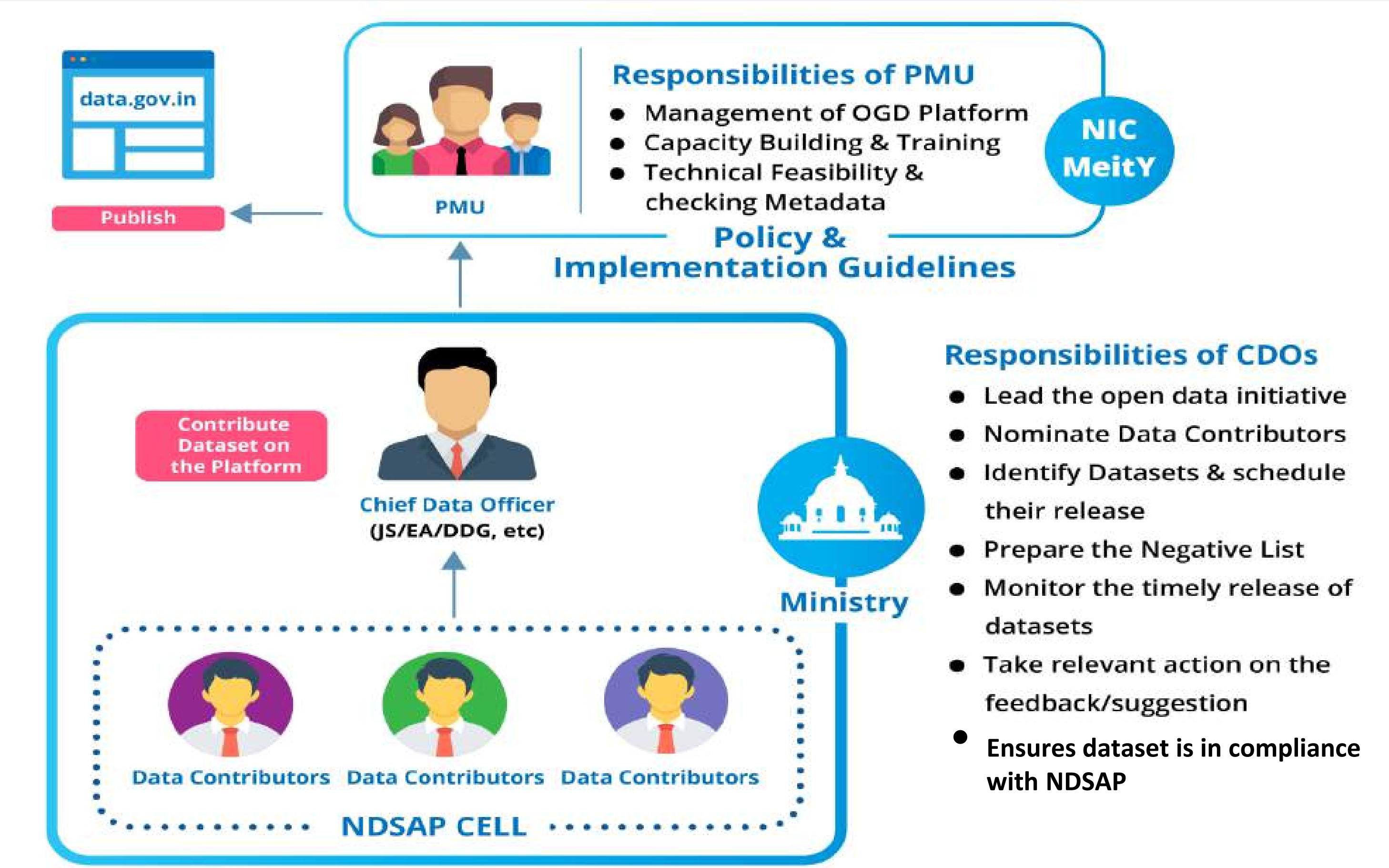
Agriculture

Education

Energy

Health

Data Governance (1/2)



data.gov.in

Publish

PMU

Responsibilities of PMU

- Management of OGD Platform
- Capacity Building & Training
- Technical Feasibility & checking Metadata

NIC
MeitY

Policy & Implementation Guidelines

Contribute Dataset on the Platform

Chief Data Officer
(JS/EA/DDG, etc)

Ministry

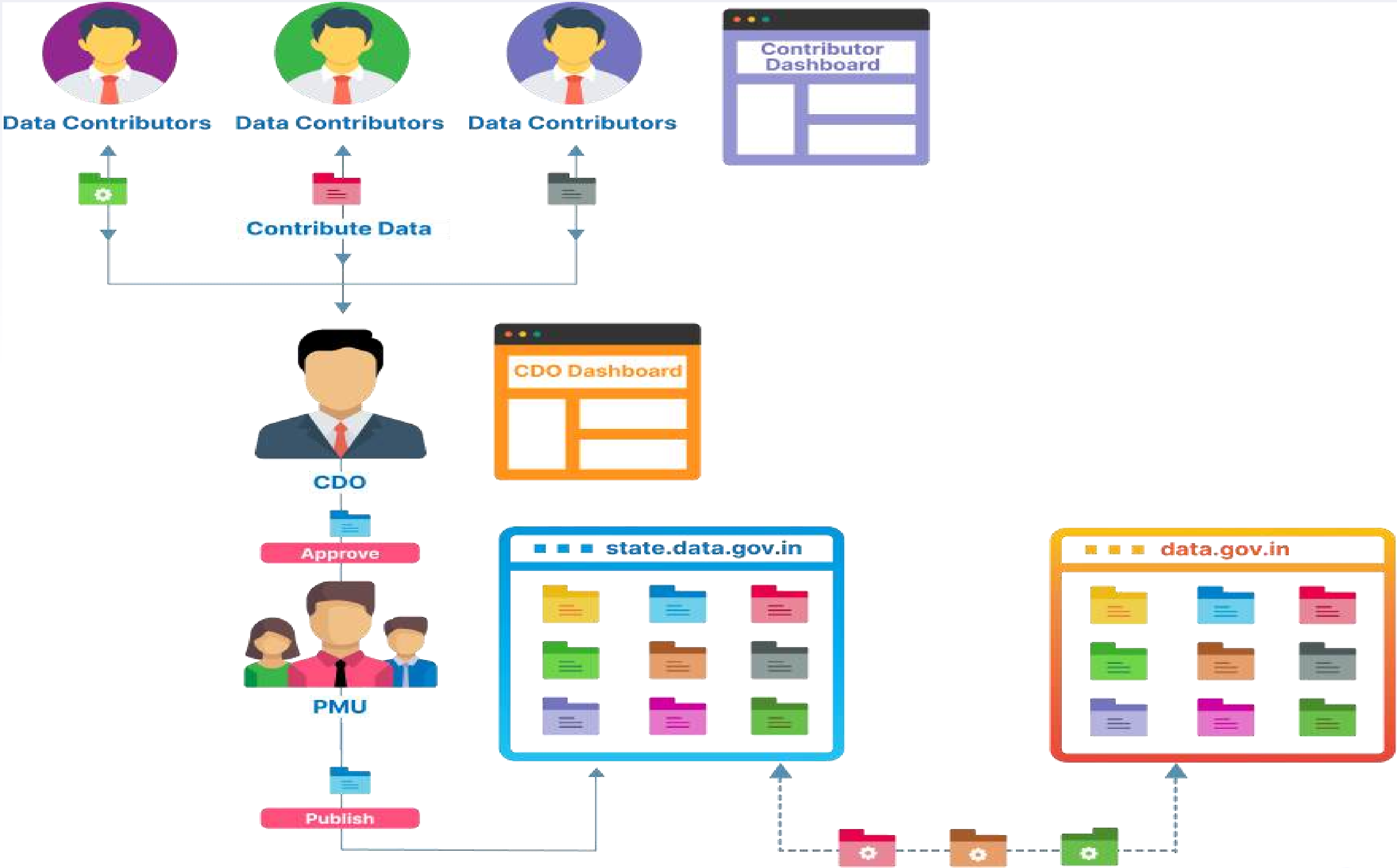
Data Contributors Data Contributors Data Contributors

NDSAP CELL

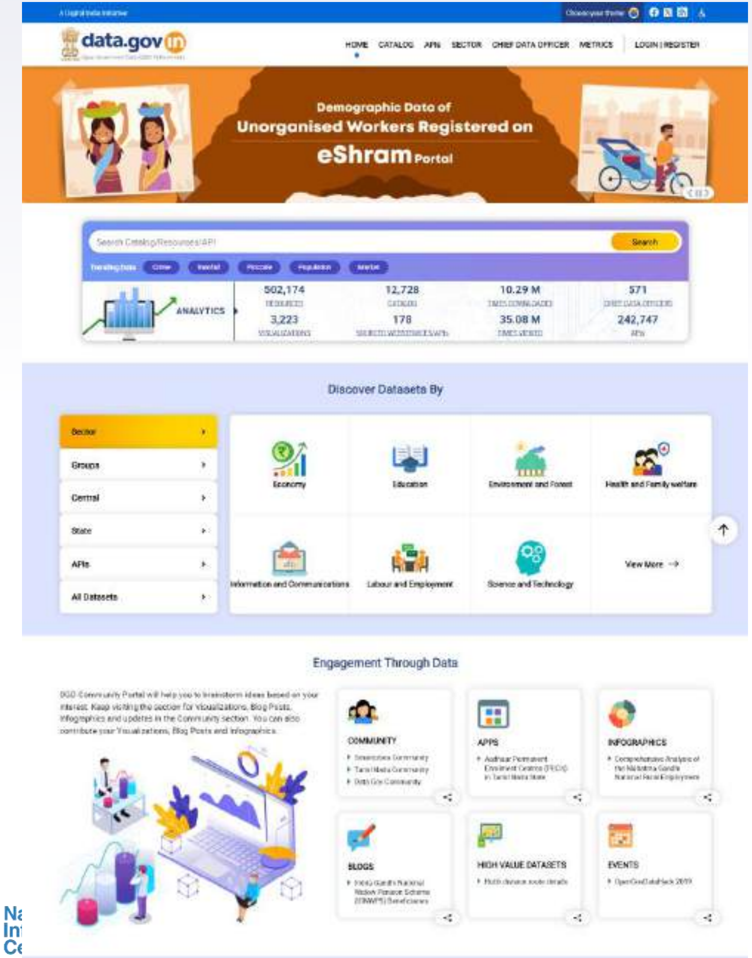
Responsibilities of CDOs

- Lead the open data initiative
- Nominate Data Contributors
- Identify Datasets & schedule their release
- Prepare the Negative List
- Monitor the timely release of datasets
- Take relevant action on the feedback/suggestion
- **Ensures dataset is in compliance with NDSAP**

Data Governance (2/2)


Data Contributors
Data Contributors
Data Contributors
Contributor Dashboard
Contribute Data
CDO
CDO Dashboard
Approve
PMU
Publish
state.data.gov.in
data.gov.in

OGD Platform - https://data.gov.in



data.gov.in

HOME CATALOG APIs SECTOR CHIEF DATA OFFICER METRICS LOGIN | REGISTER

Demographic Data of Unorganised Workers Registered on eShram Portal

Search Catalog/Resources/API

Search

ANALYTICS

502,174	12,728	10.29 M	571
RESOURCES	CATALOG	TIMES DOWNLOADED	CHIEF DATA OFFICERS
3,223	178	35.08 M	242,747
VISUALIZATIONS		TIMES VIEWED	APIs

Discover Datasets By

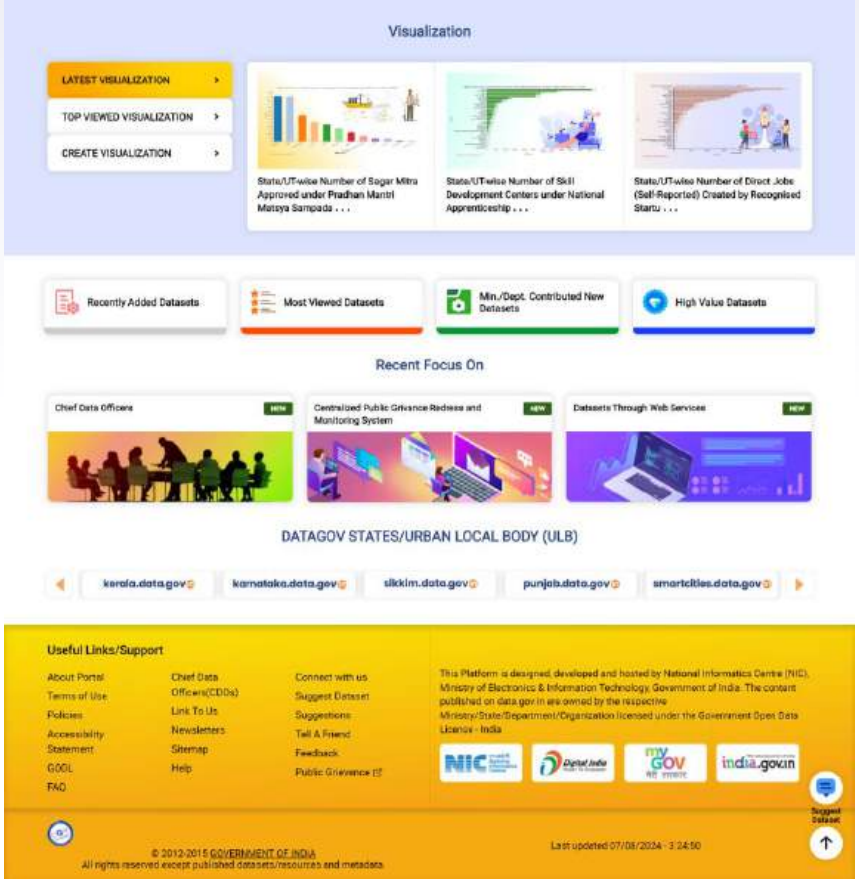
- Sector
- Groups
- Central
- State
- APIs
- All Datasets

Economy · Education · Environment and Forest · Health and Family welfare · Information and Communications · Labour and Employment · Science and Technology · View More →

Engagement Through Data

OGD Community Portal will help you to brainstorm ideas based on your interest. Keep visiting the section for Visualizations, Blog Posts, Infographics and updates in the Community section. You can also contribute your Visualizations, Blog Posts and Infographics.

COMMUNITY · APPS · INFOGRAPHICS · BLOGS · HIGH VALUE DATASETS · EVENTS



Visualization

- LATEST VISUALIZATION
- TOP VIEWED VISUALIZATION
- CREATE VISUALIZATION

State/UT-wise Number of Sagar Mitra Approved under Pradhan Mantri Matsya Sampada . . .

State/UT-wise Number of Skill Development Centers under National Apprenticeship . . .

State/UT-wise Number of Direct Jobs (Self-Reported) Created by Recognised Startu . . .

Recently Added Datasets · Most Viewed Datasets · Min./Dept. Contributed New Datasets · High Value Datasets

Recent Focus On

Chief Data Officers NEW

Centralized Public Grievance Redress and Monitoring System NEW

Datasets Through Web Services NEW

DATAGOV STATES/URBAN LOCAL BODY (ULB)

kerala.data.gov · karnataka.data.gov · sikkim.data.gov · punjab.data.gov · smartcities.data.gov

Useful Links/Support

About Portal	Chief Data Officers(CDOs)	Connect with us
Terms of Use	Link To Us	Suggest Dataset
Policies	Newsletters	Suggestions
Accessibility Statement	Sitemap	Tell A Friend
GODL	Help	Feedback
FAQ		Public Grievance

This Platform is designed, developed and hosted by National Informatics Centre (NIC), Ministry of Electronics & Information Technology, Government of India. The content published on data.gov.in are owned by the respective Ministry/State/Department/Organization licensed under the Government Open Data Licence - India

© 2012-2015 GOVERNMENT OF INDIA
All rights reserved except published datasets/resources and metadata

Last updated 07/08/2024 - 3:24:50

OGD Platform Features (1/2)

Datasets in Machine Readable Formats

Standard Metadata Elements

Filter by Sector, Ministry, State, Jurisdiction, most relevant, etc.

User friendly Search Engine with popular & advanced features

Social Share, Widgets to Embed Dataset and Data Release Alert Service

CDOs can contribute Data Files, APIs & Data URLs

On the fly auto Conversion from One format to Other

Visualization Engine

High Value Data can also be marked

OGD Platform Features (2/2)

Resources (Datasets/Apps) Catalog as Service

Bulk Download of a Catalog in Zip

Preview of Datasets

OGD Community Portal for Apps, Blogs, Visualization, Info-graphics

Event Portal for OGD Events

Widget to Consume/Link filtered set of catalogs (embed)

Subscribe a Catalog

Suggest/Endorse Datasets

Datasets APIs as a Service

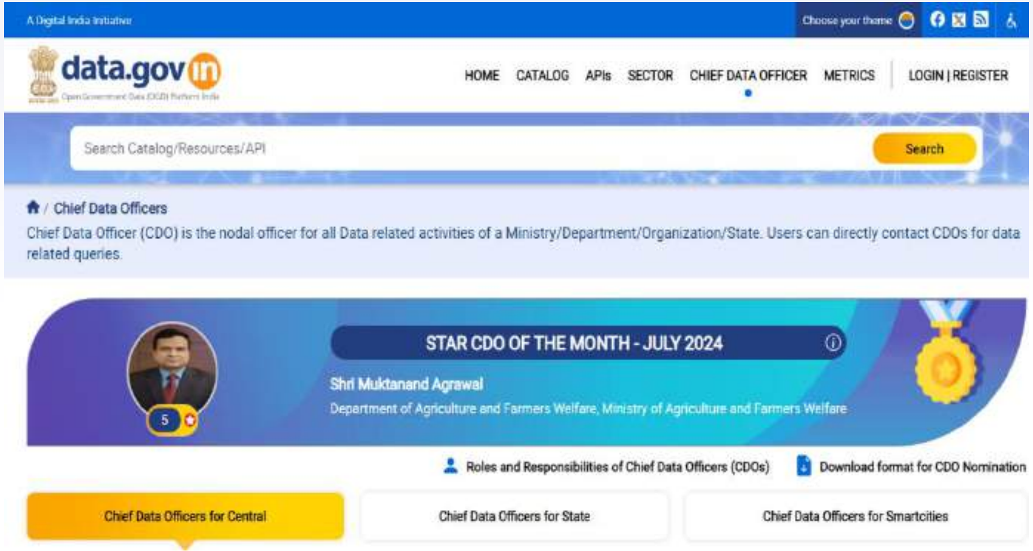
State Dashboard & Software as a Service (SaaS)

CDO and Contributor Dashboard

Analytics & Metrics

Star CDO of the Month

- No. of Datasets Published in Previous Month
- No. of Datasets for which data coming through APIs
- No. of Downloads
- User Rating



A Digital India Initiative

Choose your theme

data.gov.in

Open Government Data (OGD) Platform India

HOME CATALOG APIs SECTOR CHIEF DATA OFFICER METRICS LOGIN | REGISTER

Search Catalog/Resources/API

Search

/ Chief Data Officers

Chief Data Officer (CDO) is the nodal officer for all Data related activities of a Ministry/Department/Organization/State. Users can directly contact CDOs for data related queries.

STAR CDO OF THE MONTH - JULY 2024

Shri Muktanand Agrawal

Department of Agriculture and Farmers Welfare, Ministry of Agriculture and Farmers Welfare

5

Roles and Responsibilities of Chief Data Officers (CDOs)

Download format for CDO Nomination

Chief Data Officers for Central

Chief Data Officers for State

Chief Data Officers for Smartcities

OGD in Numbers

471 nominated Chief Data Officers form Central and State Ministries/Depts

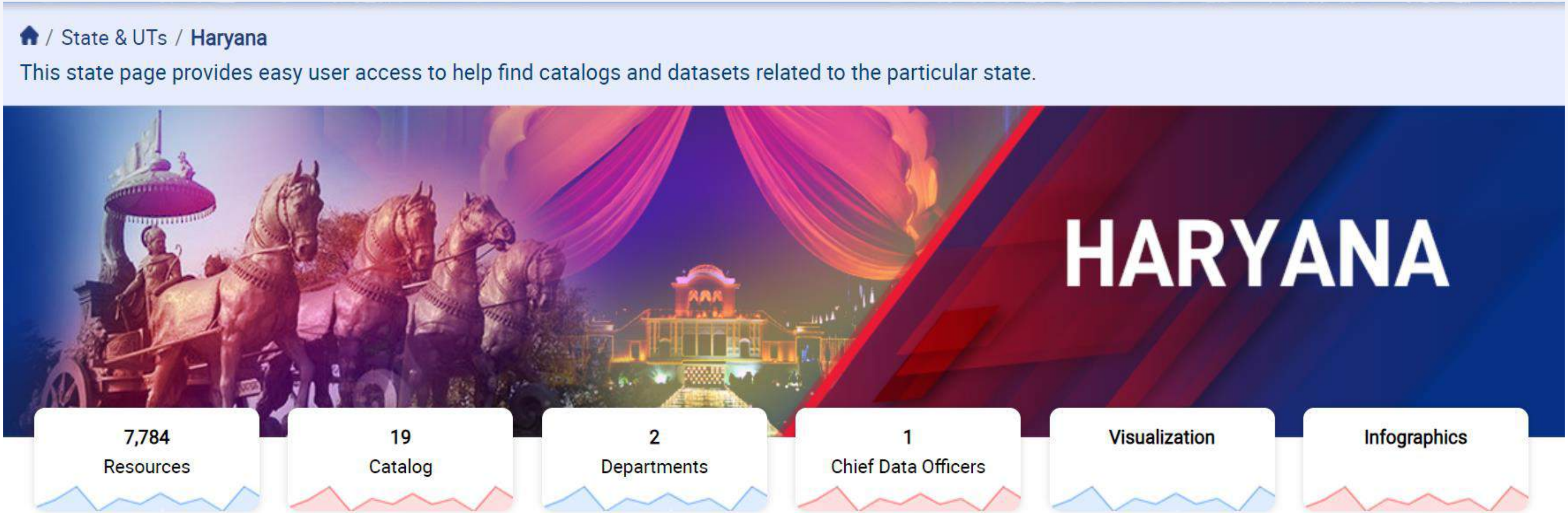
5.02 Lakh+ Datasets under 12,728 catalogs
More than 35% Datasets from Web-Services/APIs

Datasets 35.08 Million times Viewed & 10.29 Million times Downloaded

2,42,747+ Data APIs Created viz. eSharam, KCC, Company Master, HMIS, Mandi Price, Company Master, Air Quality Index, etc.

3,223+ Visualizations Created

State Dashboards & SaaS Instances (1/3)



/ State & UTs / **Haryana**

This state page provides easy user access to help find catalogs and datasets related to the particular state.

7,784
Resources

19
Catalog

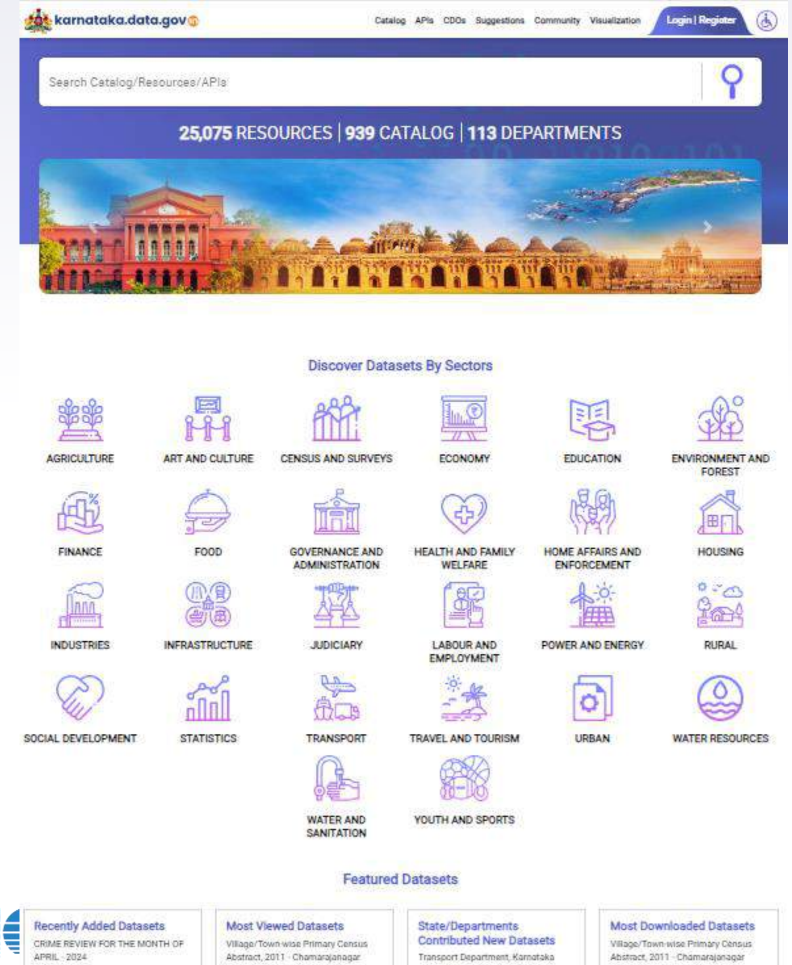
2
Departments

1
Chief Data Officers

Visualization

Infographics

State Dashboards & SaaS Instances (2/3)


karnataka.data.gov.in
Catalog
APIs
CDOs
Suggestions
Community
Visualization
Login | Register
Search Catalog/Resources/APIs

25,075 RESOURCES | **939** CATALOG | **113** DEPARTMENTS

Discover Datasets By Sectors

- AGRICULTURE
- ART AND CULTURE
- CENSUS AND SURVEYS
- ECONOMY
- EDUCATION
- ENVIRONMENT AND FOREST
- FINANCE
- FOOD
- GOVERNANCE AND ADMINISTRATION
- HEALTH AND FAMILY WELFARE
- HOME AFFAIRS AND ENFORCEMENT
- HOUSING
- INDUSTRIES
- INFRASTRUCTURE
- JUDICIARY
- LABOUR AND EMPLOYMENT
- POWER AND ENERGY
- RURAL
- SOCIAL DEVELOPMENT
- STATISTICS
- TRANSPORT
- TRAVEL AND TOURISM
- URBAN
- WATER RESOURCES
- WATER AND SANITATION
- YOUTH AND SPORTS

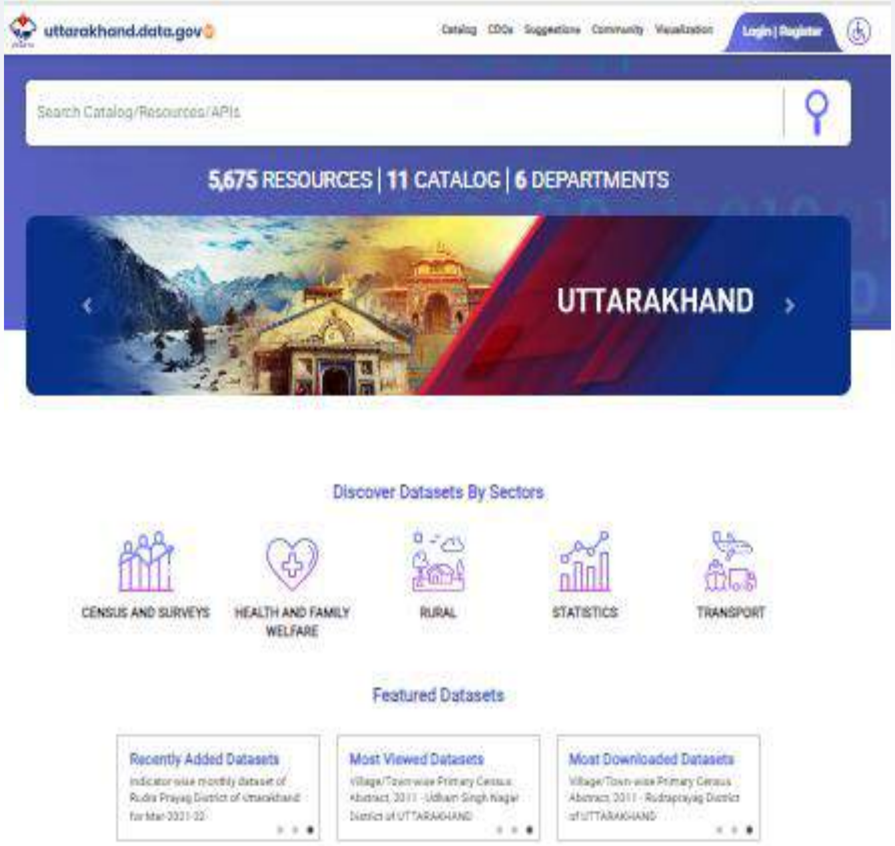
Featured Datasets

Recently Added Datasets
CRIME REVIEW FOR THE MONTH OF APRIL - 2024

Most Viewed Datasets
Village/Town-wise Primary Census Abstract, 2011 - Chamarajanagar

State/Departments Contributed New Datasets
Transport Department, Karnataka

Most Downloaded Datasets
Village/Town-wise Primary Census Abstract, 2011 - Chamarajanagar


uttarakhand.data.gov.in
Catalog
CDOs
Suggestions
Community
Visualization
Login | Register
Search Catalog/Resources/APIs
UTTARAKHAND

5,675 RESOURCES | **11** CATALOG | **6** DEPARTMENTS

Discover Datasets By Sectors

- CENSUS AND SURVEYS
- HEALTH AND FAMILY WELFARE
- RURAL
- STATISTICS
- TRANSPORT

Featured Datasets

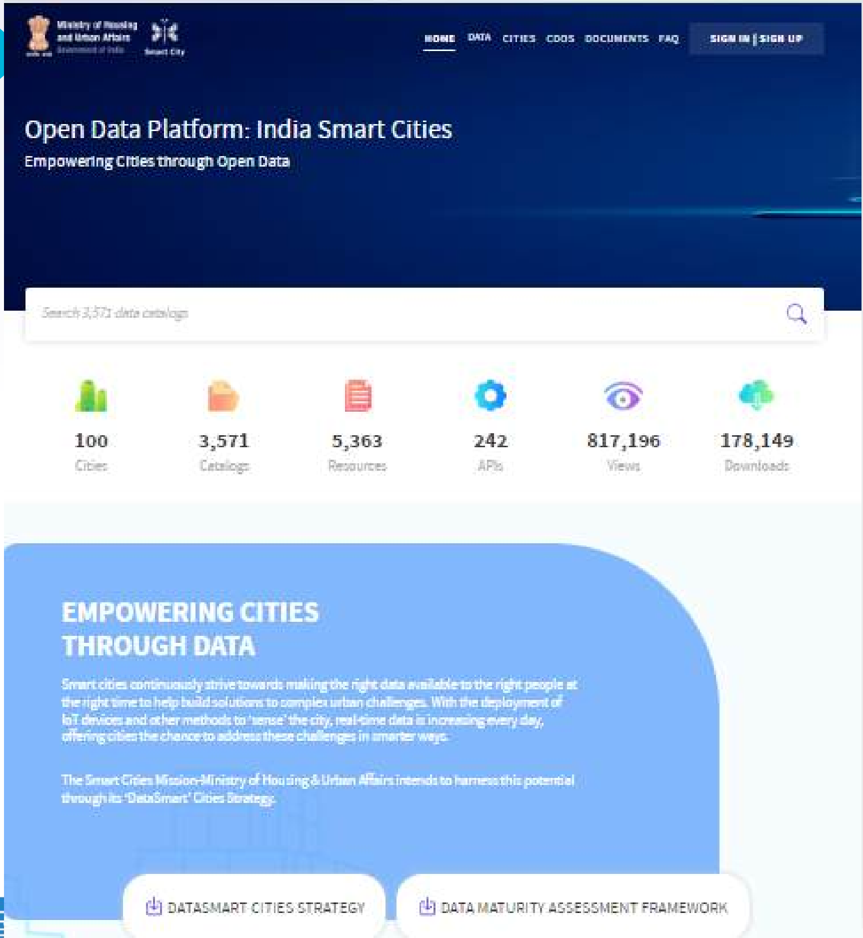
Recently Added Datasets
Indicator wise monthly dataset of Rudra Prayag District of Uttarakhand for Mar-2021-22

Most Viewed Datasets
Village/Town-wise Primary Census Abstract, 2011 - Udham Singh Nagar District of UTTARAKHAND

Most Downloaded Datasets
Village/Town-wise Primary Census Abstract, 2011 - Rudraprayag District of UTTARAKHAND

data.gov.in
Open Government Data (OGD) Platform India

State Dashboards & SaaS Instances (3/3)


Ministry of Housing and Urban Affairs
Government of India
Smart City
HOME
DATA
CITIES
CDOS
DOCUMENTS
FAQ
SIGN IN | SIGN UP

Open Data Platform: India Smart Cities

Empowering Cities through Open Data

Search 3,571 data catalogs

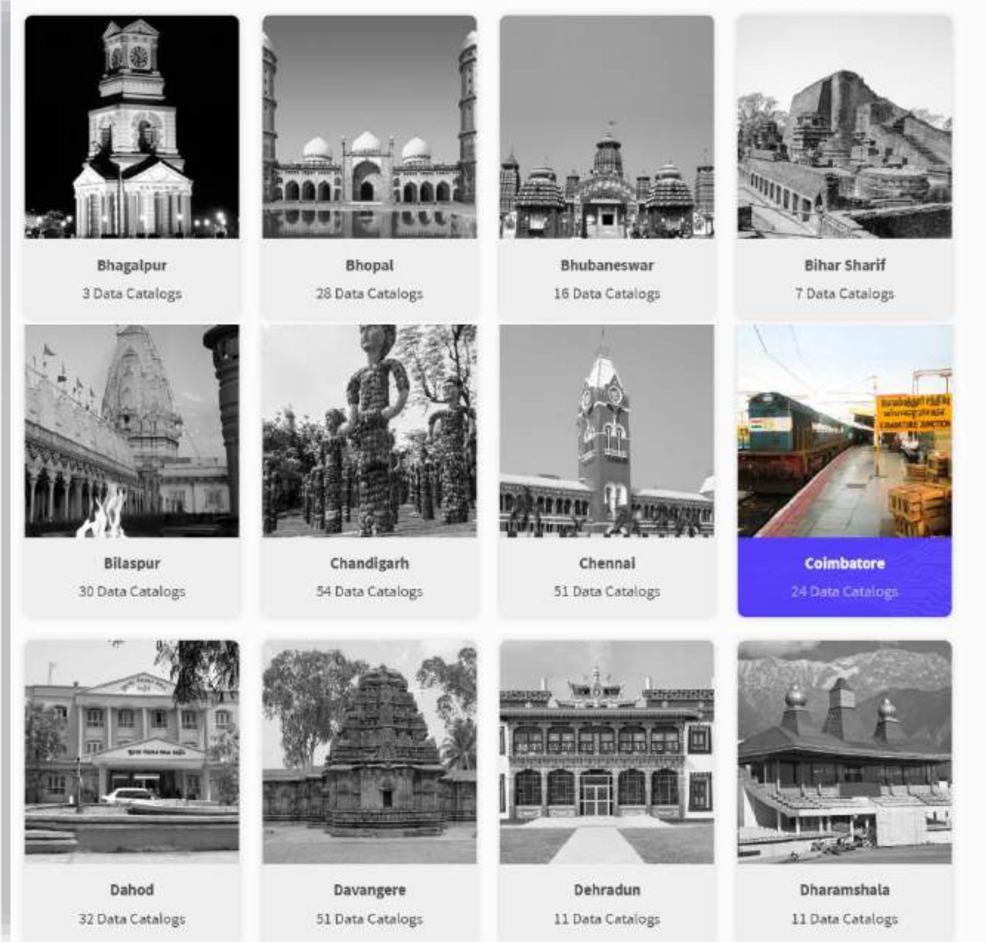
100	3,571	5,363	242	817,196	178,149
Cities	Catalogs	Resources	APIs	Views	Downloads

EMPOWERING CITIES THROUGH DATA

Smart cities continuously strive towards making the right data available to the right people at the right time to help build solutions to complex urban challenges. With the deployment of IoT devices and other methods to 'sense' the city, real-time data is increasing every day, offering cities the chance to address these challenges in smarter ways.

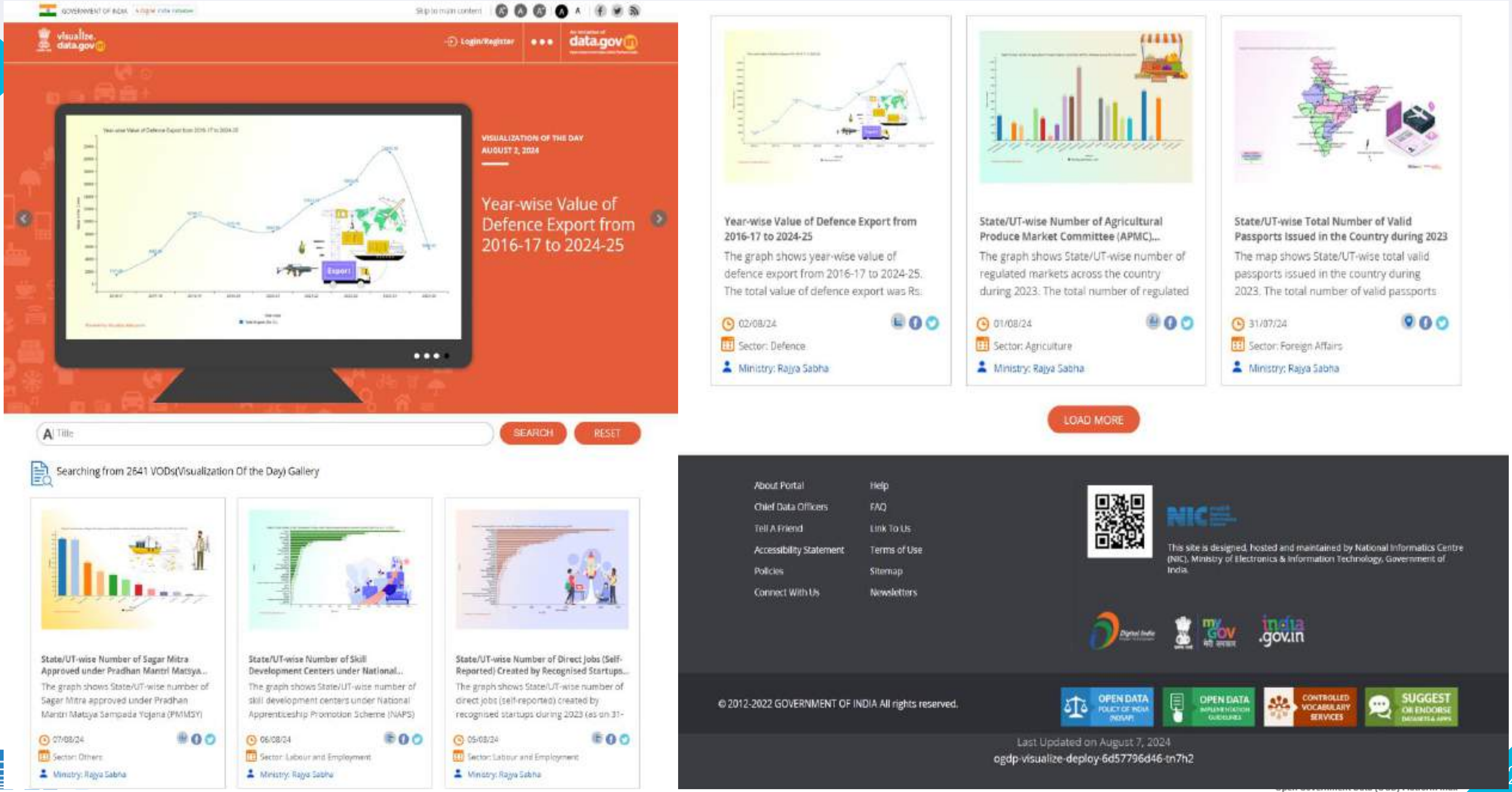
The Smart Cities Mission-Ministry of Housing & Urban Affairs intends to harness this potential through its 'DataSmart' Cities Strategy.

DATASMART CITIES STRATEGY
DATA MATURITY ASSESSMENT FRAMEWORK



City	Data Catalogs
Bhagalpur	3 Data Catalogs
Bhopal	28 Data Catalogs
Bhubaneswar	16 Data Catalogs
Bihar Sharif	7 Data Catalogs
Bilaspur	30 Data Catalogs
Chandigarh	54 Data Catalogs
Chennai	51 Data Catalogs
Coimbatore	24 Data Catalogs
Dahod	32 Data Catalogs
Davangere	51 Data Catalogs
Dehradun	11 Data Catalogs
Dharamshala	11 Data Catalogs

Visualization Engine - https://visualize.data.gov.in


GOVERNMENT OF INDIA
Skip to main content
visualize.data.gov.in
Login/Register
data.gov.in

VISUALIZATION OF THE DAY
AUGUST 2, 2024

Year-wise Value of Defence Export from 2016-17 to 2024-25

Year-wise Value of Defence Export from 2016-17 to 2024-25
Export

Title

SEARCH RESET

Searching from 2641 VODs(Visualization Of the Day) Gallery

State/UT-wise Number of Sagar Mitra Approved under Pradhan Mantri Matsya...
The graph shows State/UT-wise number of Sagar Mitra approved under Pradhan Mantri Matsya Sampada Yojana (PMMSY)
07/08/24
Sector: Others
Ministry: Rajya Sabha

State/UT-wise Number of Skill Development Centers under National...
The graph shows State/UT-wise number of skill development centers under National Apprenticeship Promotion Scheme (NAPS)
06/08/24
Sector: Labour and Employment
Ministry: Rajya Sabha

State/UT-wise Number of Direct Jobs (Self-Reported) Created by Recognised Startups...
The graph shows State/UT-wise number of direct jobs (self-reported) created by recognised startups during 2023 (as on 31-
05/08/24
Sector: Labour and Employment
Ministry: Rajya Sabha

Year-wise Value of Defence Export from 2016-17 to 2024-25
The graph shows year-wise value of defence export from 2016-17 to 2024-25. The total value of defence export was Rs.
02/08/24
Sector: Defence
Ministry: Rajya Sabha

State/UT-wise Number of Agricultural Produce Market Committee (APMC)...
The graph shows State/UT-wise number of regulated markets across the country during 2023. The total number of regulated
01/08/24
Sector: Agriculture
Ministry: Rajya Sabha

State/UT-wise Total Number of Valid Passports Issued in the Country during 2023
The map shows State/UT-wise total valid passports issued in the country during 2023. The total number of valid passports
31/07/24
Sector: Foreign Affairs
Ministry: Rajya Sabha

LOAD MORE

About Portal
Chief Data Officers
Tell A Friend
Accessibility Statement
Policies
Connect With Us

Help
FAQ
Link To Us
Terms of Use
Sitemap
Newsletters

This site is designed, hosted and maintained by National Informatics Centre (NIC), Ministry of Electronics & Information Technology, Government of India.

© 2012-2022 GOVERNMENT OF INDIA All rights reserved.

OPEN DATA POLICY OF INDIA (NDSAP)
OPEN DATA IMPLEMENTATION GUIDELINES
CONTROLLED VOCABULARY SERVICES
SUGGEST OR ENDORSE DATASETS & APPS

Last Updated on August 7, 2024
ogdp-visualize-deploy-6d57796d46-tn7h2

Community Engagement & Events

#OpenGovDataHack-2019
Multi-city Coding & Hack Challenge For Data Driven Innovation By Startups, Entrepreneurs and MSMEs
PRE-CHALLENGE SENSITISATION WORKSHOP
19th August
PRE-HACKATHON SENSITISATION WORKSHOP
RK UNIVERSITY

How OGD Platform Leverage Data Ecosystem

Monitored by PMO for G2G Data Sharing

Principal Scientific Adviser on Open Access of Scientific Data

Chief Economic Advisor focused on economic Data Sharing

Help States to Adopt/Draft State Data Policies

Tech Support to States and ULBs for create own Data Portal Instances

Develop and Support Smart Cities Mission for Open Data Portal

Encouraging Machine to Machine Data Sharing

Data Driven Innovation and Decision Making

Help in International Ranking e.g. UN eGov, Open Data Barometer, ROLI, etc.

Data Governance Quality Index (DGQI) helps in encouraging Data by Default

Part of all International Treaties

Support by Min of Finance to States for release of Open Data

OGD Used for Major Product & Services

CLIMATE FIELDVIEW

THE HUFFINGTON POST

BreeZo

socialcops

opencorporates

How INDIA Lives BETA

Blue Sky Analytics

fusionbase

agribazaar

BOOM

Animall

iS. IndiaSpend

opencontractingindia
অসম I Assam

THE HINDU

civic data lab

Justice Hub

OpenBudgetsIndia BETA

CSIRO

SUSTAINABLE DEVELOPMENT GOALS
17 GOALS TO TRANSFORM OUR WORLD
NITI Aayog

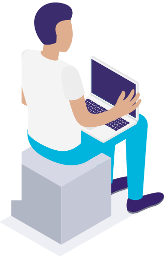
cropdata
Connecting Farmers

THANK YOU

Write to us at ndsap@gov.in

OGD Brochure

NDSAP



GODL

National Informatics Centre

data.gov.in

Open Government Data (OGD) Platform India