ASPD/Oct/2024/2

Government of India
Ministry of Statistics and Programme Implementation

Data Dissemination: National Metadata Structure (NMDS)

October 2024

Data Dissemination: National Metadata Structure (NMDS)

(2nd Edition)

October 2024

<div style="text-align:center">

Foreword

</div>

Standardisation is one of the basic pre-requisites for any statistical production process. It helps in making statistical processes more efficient and robust, and to improve reporting on the quality of processes and output. Standards support the sharing of knowledge-based experiences and methodologies and, consequently, the sharing of tools, data, services and resources. The implementation of standard methods and tools, in turn, improves the comparability of statistical outputs, thus benefiting users of statistics.

As a standard, this publication encourages sharing information to Indian official statistics and the spread of best practices, in particular for quality and metadata reporting. It aims to promote harmonised quality reporting across National Statistical System (NSS), and, thus, to facilitate cross-comparisons of processes and outputs. The Handbook applies to all the stakeholders of NSS and others in their roles as producers, compilers and disseminators of Indian official statistics.

In its first, Ministry of Statistics & Programme Implementation (MoSPI), after due consultation with the stakeholders of the NSS, notified first edition of National Metadata Structure (NMDS) in August 2021. While reporting compliance on various elements of NMDS in respect of statistical products of MoSPI, it emerged that there is a need to further distinguish the elements of NMDS for surveys/census, administrative records/statistics, and other datasets/indices in terms of their appropriate applicability. Further, it was also felt that various elements of NMDS may also be classified as mandatory and optional for reporting purposes. This arrangement is likely to bring to the fore improved dissemination of statistical products on various elements of NMDS.

ASPD/Oct/2024/2

The publication has been prepared by the Administrative Statistics and Policy Division (ASPD) under the guidance of ADG (ASPD) and DG (DG), with extensive inputs from various stakeholders of the NSS.

I Congratulate the team ASPD led by DDG (ASPD) as well as all the other colleagues in the NSO who have helped to prepare this publication.

Secretary, MoSPI

New Delhi
October, 2024

ASPD/Oct/2024/2

Officials of ASPD associated with the publication

|	
Dr. Praveen Shukla	Additional Director General
Dr. Pankaj K. P. Shreyaskar	Deputy Director General
Ms. Aqsa Ilahi	Joint Director
Ms. Renu Verma	Deputy Director
Ms. Himanchali Rana	Deputy Director
Sh. Sanjoy Roy	Assistant Director
Sh. Hemant Kumar	Senior Statistical Officer
Sh. Mohd Azharuddin	Junior Statistical Officer
Sh. Ankit Kumar	Junior Statistical Officer

Introduction

National Statistics 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[1] 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.
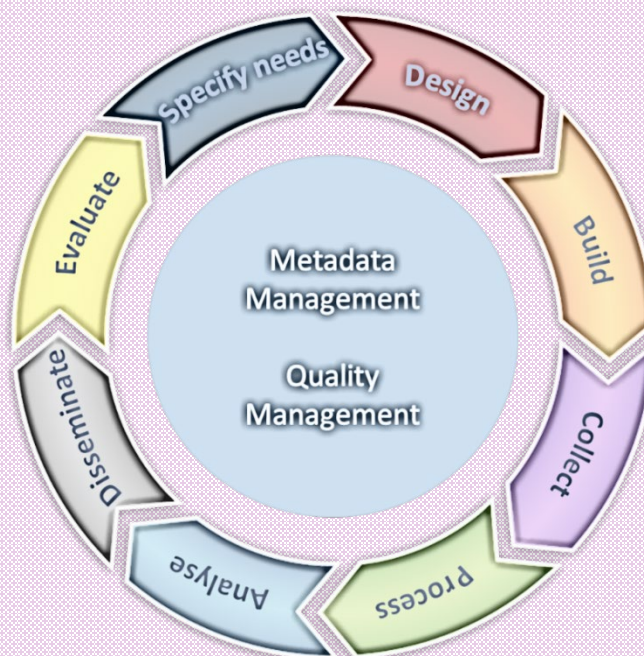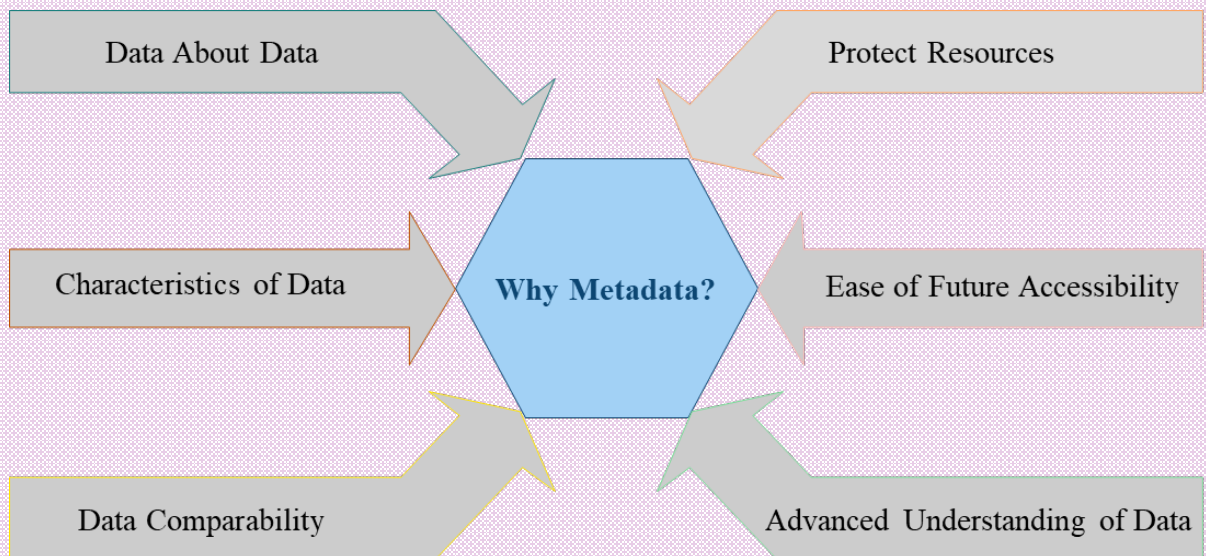

The Metadata process is fully integrated in the Generic Statistical Business Process Model[2] (GSBPM) which has metadata as one of the key elements in the version 5.1.

---

[1] The term dataset (also termed as data set) refers to a file that contains one or more records. Any named group of records is called a dataset. Datasets can hold information such as medical records or insurance records, to be used by a program running on the system. Datasets are also used to store information needed by applications or the operating system itself, such as source programs, macro libraries, or system variables or parameters. For datasets that contain readable text, one can print them or display them on a console (many data sets contain load modules or other binary data that is not really printable). Datasets can be catalogued, which permits the data set to be referred to by name without specifying where it is stored. (Source- IBM; https://developer.ibm.com/data/)

[2] 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. Statistical Data and Metadata

C.1. To avoid confusion about which "data" and "metadata" are the intended content of the formats specified here, a statement of scope is offered. Statistical "data" are sets of often numeric observations which typically have time associated with them. They are associated with a set of metadata values, representing specific concepts, which act as identifiers and descriptors of the data. These metadata values and concepts can be understood as the named dimensions of a multi-dimensional co-ordinate system, describing what is often called a "cube" of data.

C.2. The term "metadata" is very broad indeed. A distinction can be made between "structural" metadata – those concepts used in the description and identification of statistical data and metadata – and "reference" metadata – the larger set of concepts that describe and qualify statistical data sets and processing more generally, and which are often associated not with specific observations or series of data, but with entire collections of data or even the institutions which provide that data[3].

---

[3] SDMX Standards (2020): Framework for SDMX Technical Standards version 2.1

D. Types of Metadata

D.1. According to terminology agreed to describe types of statistical metadata[4], there are two types of metadata: *structural metadata* and *reference metadata.*

Structural metadata identify and describe data, and reference metadata. Structural metadata are needed to identify, use, and process data matrixes and data cubes, e.g. names of columns or Dimensions of statistical cubes. Structural metadata must be associated with the statistical data and reference metadata, otherwise it becomes impossible to identify, retrieve and navigate the data or reference metadata.

In SDMX structural metadata are not limited to describing the structure of data and reference metadata. The structural metadata in SDMX include many of the other constructs to be found in the SDMX Information Model including data discovery, data and metadata Constraints (used for both data validation and data discovery), data and structure mapping, data and metadata reporting, statistical processes.[5]

Reference metadata describe the contents and quality of the statistical data. Preferably, reference metadata should include all of the following: a) "conceptual" metadata, describing the concepts used and their practical implementation, allowing users to understand what the statistics are measuring and, thus, their fitness for use; b) "methodological" metadata, describing methods used for the generation of the data (e.g. sampling, collection methods, editing processes); and c) "quality" metadata, describing the different quality dimensions of the resulting statistics (e.g. timeliness, accuracy).[6]

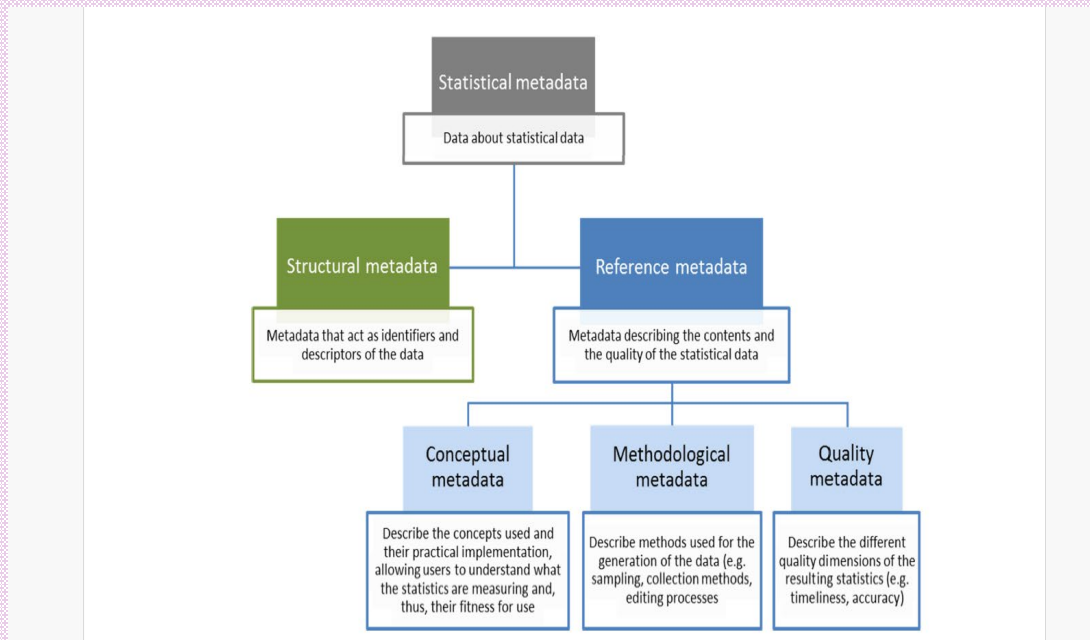The pictorial presentation of the metadata ecosystem may be presented as under:

---

[4]SDMX Guidelines (2020): SDMX Glossary version 2.1

[5] Paragraph 3.3 of SDMX Standards (2020): Framework for SDMX Technical Standards version 2.1

[6] Ibid

Figure: **Types of statistical metadata**[7]

This publication is mainly concerned with reference metadata and how they are presented to users when data are published.

There must be flexibility in the presentation of metadata, so that statistical products have the level of detail appropriate for the target audience(s). This document recommends metadata that are mandatory (always required); and optional (data producers are encouraged for dissemination but are not mandatorily required)[8] for the datasets produced through various sources such as census, survey, administrative data, alternative data, as also for the indices, etc.

---

[7] United Nations Economic Commission for Europe (2013): Getting the Facts Right-A guide to presenting metadata drawn based on SDMX Guidelines

[8] Metadata registries standard of ISO 11179 provides for mandatory, conditional and optional classifications. Until the referential metadata reaches its maturity level, the classification has only been relied as mandatory and optional.

E. Metadata standards

E.1. National and international statistical organizations face common challenges in managing their metadata, and there has been significant effort to develop shared metadata standards and models. Some of the metadata standards which are widely followed are presented hereunder:

E.2. Statistical Data and Metadata Exchange (SDMX)[9]: The Statistical Data and Metadata Exchange (SDMX) initiative sets technical standards and content-oriented guidelines to facilitate the exchange of statistical data and metadata. Used by a large number of international and national statistical organizations, SDMX is maintained by a group of seven sponsors: The Bank for International Settlements, the European Central Bank, Eurostat, the International Monetary Fund, the Organisation for Economic Cooperation and Development (OECD), the United Nations and the World Bank.

The introductory document (https://sdmx.org/wp-content/uploads/SDMX_COG_2016_Introduction.pdf) provides a general description of the various components of the SDMX Content-Oriented Guidelines, namely: cross-domain concepts, code lists, subject-matter domains, glossary, and implementation-specific guidelines. This material is made available to users to enhance and make more efficient the exchange of statistical data and metadata using the SDMX standard.

The SDMX Content-Oriented Guidelines facilitate mutual understanding of the content of the SDMX data and metadata files through the use of common statistical concepts in the underlying metadata.

SDMX regularly reports on its activities to the UN Statistical Commission (UNSC). At the 39th Session of the Commission in 2008, SDMX was recognized as the preferred standard for the exchange and sharing of data and metadata. The Commission also encouraged implementation by national and international statistical organizations.

E.3. Data Documentation Initiative (DDI)[10]: The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational

---

[9] https://sdmx.org/

[10] https://ddialliance.org/

methods in the social, behavioral, economic, and health sciences. DDI is a free standard that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving. Documenting data with DDI facilitates understanding, interpretation, and use -- by people, software systems, and computer networks.

There are three different DDI (https://ddialliance.org/) standards:

- DDI Codebook (DDI-C)
  - Intended to document simple or single survey data
  - Has the same information describing a dataset that would be available in the codebook for a dataset
  - Used in the Dataverse data repository
  - Based on XML

- DDI Lifecycle (DDI-L)
  - Can be used to document each stage of the lifecycle of a dataset
  - Can be used for longitudinal datasets
  - Can describe more than one dataset
  - Based on XML

- DDI Cross-Domain Integration (DDI-CDI)
  - Suitable for integration of data from different research disciplines or domains (cross-domain) purposes
  - Focuses on data description and process/provenance
  - Handles data from different structure types (wide, long multidimensional, key value/big data)
  - Based on UML (Unified Modeling Language)
  - Can be expressed in different syntax representations, for example XML

The DDI updates their resource documents regularly and incorporates into its resource materials contemporary developments in the field.

E.4. Metadata registries ISO (International Organization for Standardization)/IEC (International Electrotechnical Commission) 11179[11]- This is a standard for describing and managing the meaning and representation of data. The basic semantic unit is a concept. Both DDI and SDMX are based on ISO/IEC 11179 for their descriptions of data and their use of concepts as a basic semantic unit.

ISO/IEC 11179 specifies the structure of a Metadata registry and provides a metamodel for registry common facilities. This metamodel is intended to be extended by other parts of ISO/IEC 11179 for specific purposes (https://www.iso.org/standard/78914.html).

This standardization and registration allow for easier locating, retrieving, and transmitting data from disparate databases. The standard defines the how metadata are conceptually modelled and how they are shared among parties, but does not define how data is physically represented as bits and bytes.

E.5. Metadata Common Vocabulary (MCV)[12]- The Metadata Common Vocabulary (MCV) is an SDMX repository which contains concepts (and related definitions) to which terminology used in structural and reference metadata of international organisations and national data producing agencies may be mapped. The MCV covers a selected range of metadata concepts: (1) General metadata concepts, mostly derived from ISO, UNECE and UN documents, useful for providing a general context to metadata management;

(2) Metadata terms describing statistical methodologies (frequency, reference period, data collection, source, adjustment, etc.);

(3) Metadata for assessing quality (accuracy, timeliness, etc.), and

(4) Terms referring specifically to data and metadata exchange (terminology from the SDMX information model and from existing data structure definitions, etc.).

---

[11] https://www.iso.org/obp/ui/#iso:std:iso-iec:11179:-1:ed-4:v1:en

[12] https://sdmx.org/wp-content/uploads/Content_04_Draft_Guidelines_Metadata_Common_Vocabulary-MARCH-2006-1.pdf

The MCV is not intended to cover the whole range of statistical terminology, as this area is already covered by other general and domain-specific glossaries. The focus of the MCV is largely those terms that are normally used for building and understanding metadata systems. A metadata glossary is necessarily linked to a series of other subject-specific glossaries (on classifications, on data editing, on subject-matter statistical areas) or to more universal statistical glossaries such as Eurostat's CODED or the OECD Glossary of Statistical Terms. These more extensive glossaries also contain numerous terms and definitions relevant to specific statistical domains (such as prices, national accounts, external trade, etc.). The insertion within the MCV of some definitions derived from other glossaries should not be seen as a redundancy, but as a means of resolving the complex and interdisciplinary nature of metadata.

E.6. Neuchâtel Terminology Model (NTM)[13]- In 2004, the Neuchâtel Group[14] issued Neuchâtel Terminology Model Classification database object types and their attributes. The main purpose of the work was to arrive at a common language and a common perception of the structure of classifications and the links between them. The present document extends the model with variables and related concepts. The discussion includes concepts like object types, statistical unit types, statistical characteristics, value domains, populations etc. The two models together claim to provide a more comprehensive description of the structure of statistical information embodied in data items.

The Terminology Model is both a terminology and a conceptual model. It defines the key concepts that are relevant for the structuring of metadata on variables and provides the conceptual framework for the development of a database organising that metadata.

---

[13] https://www.dst.dk/Site/Dst/SingleFiles/GetArchiveFile.aspx?fi=3509740098&fo=0&ext=ukraine

[14] The statistical offices of Denmark, Norway and Sweden were among the participants of the IMIM (Integrated Metainformation Management System) project of the 4th Framework Programme of the EU. The main result of the IMIM project was a software product, BRIDGE, an object-oriented system for metadata management. In June 1999, a meeting on terminology was held in Neuchâtel, Switzerland, with participants from the statistical offices of Denmark, Norway, Sweden and Switzerland and Run Software-Werkstatt (developers of the BRIDGE software) focussing on the classification database part of BRIDGE. This was the start of the "Neuchâtel group". (https://www.dst.dk/Site/Dst/SingleFiles/GetArchiveFile.aspx?fi=595717909&fo=0&ext=ukraine)

The Neuchâtel Terminology Model has a two-level structure, consisting at the first level of the object types and, on the second level, the attributes associated with each object type. Both object types and their attributes are defined by a textual description. Since the model belongs to the semantic and conceptual sphere of metadata, it does not include metadata that are related solely to the technical aspects of a variables database. The Neuchâtel Terminology Model is generally applicable and not dependent on IT software and platforms. It may be used in any context where structured information on variables and their related concepts is required.

The Neuchâtel Terminology Model is not just a vocabulary in the sense of a mere collection of terms related to the conceptual variable. It has been developed for several purposes, using a specific methodology, ordering the concepts in a two-level structure of object types and attributes. On the first level, it specifies the object types, while on the second level, it lists the attributes associated with each object type. These attributes refer both to characteristics of the object types and to relations with other object types.

In 2006 the model was extended with variables and related concepts. The discussion includes concepts like object types, statistical unit types, statistical characteristics, value domains, populations etc. These two models together claim to provide a more comprehensive description of the structure of statistical information embodied in data items.

Metadata standards, models and guidelines form a valuable basis for statistical organizations to develop their data and metadata management systems. Compliance with international standards leads to greater consistency and interoperability within the organization. It will also help to exchange and share methods and tools with other organizations, both within the national statistical system and internationally. Effectively managing metadata throughout the statistical production process is the first step in ensuring sufficient information can be provided to data users.

F. Role of Ministry of Statistics & Programme Implementation (MoSPI) in Developing Metadata Structure

F.1 Ministry of Statistics & Programme Implementation (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.

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

G. Metadata Management

G.1 The document in the hand is a set of broad, high level principles that form the guiding framework within which metadata management can operate.

G.2 Once the metadata structure is put in implementation, 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 the little stakeholders' acceptance, if the data are incomplete or is not updated regularly.

H. National Metadata Structure (NMDS) Concepts - Index

Item No	Concept
1	**Contact**
1.1	Contact organisation*
1.2	Compiling agency*
1.3	Custodian agency*
1.4	Contact details*
2	**Data Description and Presentation**
2.1	Data description*
2.2	System of Classification*
2.3	International/National Standards Classification/Codes, if followed*
2.4	Sector coverage*
2.5	Concepts and definitions*
2.6	Unit of compilation
2.7	Population coverage
2.8	Reference Period*
2.9	Duration and period of enumeration*
2.10	Sample size/Dataset size*
2.11	Data confidentiality
3	**Institutional Mandate**
3.1	Legal acts and other agreements*
3.2	Data sharing/Data dissemination
3.3	Release calendar*
N3.4	Frequency of dissemination*
3.5	Data access*
4	**Quality Management**
4.1	Documentation on methodology*
4.2	Quality documentation
4.3	Quality assurance
5	**Accuracy and Reliability**
5.1	Sampling error
5.2	Measures of reliability
6	**Timeliness**

Item No	Concept
6.1	Timeliness*
7	**Coherence and Comparability**
7.1	Comparability – over time*
7.2	Coherence
8	**Data Processing**
8.1	Source data type*
8.2	Frequency of data collection*
8.3	Mode and method of data collection
8.4	Data validation*
8.5	Data compilation
8.6	Data identifier(s)*
9	**Metadata Update**
9.1	Metadata last posted
9.2	Metadata last update

Note:

- All elements of the NMDS with (*) are mandatorily to be disseminated
- Other elements of the NMDS are optional, however, data producing Ministry/Departments/Agencies are encouraged to disseminate as much information as available with them.
- Dissemination of some of the elements of the NMDS may be applicable for the Surveys/Censuses whereas, some other elements may be applicable for the other datasets, such as administrative data, various indices, alternative data etc.
- Administrative data is the set of units and data derived from an administrative source. Administrative data collection is set of activities involved in the collection, processing, storage and dissemination of statistical data from one or more administrative sources. The equivalent of a survey but with the source of data being administrative records rather than direct contact with respondents.

(Source- OECD, IMF, ILO, Interstate Statistical Committee of the Commonwealth of Independent States, "Measuring the Non-Observed Economy: A Handbook", Second Draft, Annex 2, Glossary, Paris, 2002)

Note:

I. 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	Organisation of the contact point(s) 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	Custodian agency	Refers to an institution or agency which has responsibility of managing use, disclosure and protection of data/statistics.	Data/statistics custodians are agencies responsible for managing the use, disclosure and protection of source data used in a statistical ecosystem. Data custodians collect or authorise to collect and hold information on behalf of a data provider (defined as an individual, household, business or other organisation which supplies data either for statistical or administrative purposes). The role of data custodians may also extend to producing source data, in addition to their role as a holder of datasets.
1.4	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)

Item No	Concept name	Definition	Guidelines
			responsible for process and its outputs, provide this name also with his/her details
2	**Data Description and Presentation**	**Description of the disseminated data which can be displayed to users as tables, graphs or maps**	
2.1	Data description	Metadata element describing the main characteristics of the Data Set in an easily understandable manner, referring to the main data and indicators disseminated.	This summary description should provide an immediate understanding of the data to users (also to those who do not have a broader technical knowledge of the Data Set in question). Data can be displayed to users as tables, graphs or maps. According to the United Nations' Fundamental Principles of Official Statistics, the choice of appropriate presentation methods should be made in accordance with professional considerations. Data presentation includes the description of the Data Set disseminated with the main variables covered, the classifications and breakdowns used, the reference area, a summary information on the time period covered and, if applicable, the base period used.
2.2	System of Classification	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	International/ National Standards Classification etc.	An international statistical standard is an internationally agreed statistical macro-socio-economic output framework or a cross-functional framework.	An international statistical standard is an internationally agreed statistical macroeconomic output framework or a cross-functional framework. The framework consisting of concepts, definitions, classifications and inter-related tables for integrating broad set of statistics. The framework is internally coherent and consistent and externally consistent, to the extent possible, with other international statistical standards.

Item No	Concept name	Definition	Guidelines
		International/ National standard classification, is the primary tool for collecting and presenting internationally/n ationally comparable statistics by economic activity.	
2.4	Sector coverage	Description of sectors covered by the statistics.	This metadata element describes all applicable sectors associated with the Data Set. An example is "Production units and households" in environmental accounts.
2.5	Concepts and definitions	Characteristics of statistical observations, or variables used	Define and describe briefly the main statistical observations or variables that have been observed or derived. Indicate their types.
2.6	Unit of compilatio n	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	Population coverage	Definition of the main types of population covered by the statistics or variables	The population coverage describes the types of population covered by the statistics or variables whenever applicable.
2.8	Reference Period	Timespan or point in time to which the measured observation is intended to refer.	A reference period is the time period for which statistical results are collected or calculated and to which the data refer. The time period may be a calendar year (reference year), a fiscal year, a semester, a quarter, a month and even a day (reference date).
2.9	Duration and period of	Duration of enumeration refers to the time taken to conduct	The enumeration period refers to the specific time of year when enumeration operations took place.

Item No	Concept name	Definition	Guidelines
	enumeration	the enumeration process of all nature.	
2.10	Sample size/Dataset Size	Sample size is the number of observations or individuals included in a study or experiment.	Sample size is a key feature potentially affecting the results of all experimental studies. A larger sample size can potentially enhance the precision of estimates, leading to a narrower margin of error. In other words, the results from a larger sample will likely be closer to the true population parameter. A larger sample size can also increase the power of a statistical test.
2.11	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, 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.	The concept covers provision in law assigning responsibility to specific organisations for collection, processing, and dissemination of statistics in one or several statistical domains. In addition, non-legal measures such as formal or informal administrative arrangements employed to specific organisations for collection, processing, and dissemination of statistics in one or several statistical domains should also be described.
3.2	Data sharing/Data Dissemination	Exchange of data and/or metadata in a situation involving the use of open, freely available data formats and where process	Data Sharing- In data sharing exchange, any organisation or individual can use any counterparty's data and metadata (assuming they are permitted access to it). This model requires no bilateral agreement, but only requires that data and metadata providers and consumers adhere to the standards.

Item No	Concept name	Definition	Guidelines
		patterns are known and standard. Regular or *ad hoc* publications in which the data are made available to the public.	Dissemination-This metadata element provides references to the most important data dissemination done through paper or online publications, including a summary identification and information on availability of the publication means.
3.3	Release calendar	Schedule of release dates.	An advance release calendar is the schedule for release of data, which are publicly disseminated so as to provide prior notice of the precise release dates on which a national statistical agency, other national agency, or international organisation undertakes to release specified statistical information to the public. Such information may be provided for statistical releases in the coming week, month, quarter or year.
3.4	Frequency of dissemination	The time interval at which the statistics are disseminated over a given time period.	The frequencies with which data are released, which could be different from the frequency of data collection.
3.5	Data access	The conditions and modalities by which users can access, use and interpret data. Statistics should be easy to find and obtain, presented clearly and in a way that can be understood, and available and accessible to all users in line with open data standards.	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.	This metadata element refers to the availability of documentation related to various aspects of the data, such as methodological documents, summary notes or papers covering concepts, scope, classifications and statistical techniques.
4.2	Quality documentation	Documentation on procedures applied for quality management and quality assessment.	This metadata element is used to document the methods and standards for assessing data quality, based on standard quality criteria such as relevance, accuracy and reliability, timeliness and punctuality, accessibility and clarity, comparability, and coherence.
4.3	Quality assurance	Guidelines focusing on quality in general and dealing with quality of statistical programmes, including measures for ensuring the efficient use of resources.	This metadata element refers to all the planned and systematic activities implemented that can be demonstrated to provide confidence that the data production processes will fulfil the requirements for the statistical output. This includes the design of programmes for quality management, the description of planning process, scheduling of work, frequency of plan updates, and other organisational arrangements to support and maintain planning function.
5	**Accuracy and Reliability**	**Statistics should accurately and reliably portray reality. 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	Sampling errors are distinct from errors due to imperfect selection, bias in response or estimation, errors of observation and recording, etc. For probability sampling, the random variation due to sampling can be calculated. For non-probability sampling, random errors cannot be calculated without reference to some kind of model. The totality of sampling errors in all possible samples

Item No	Concept name	Definition	Guidelines
		that only a subset of the population is enumerated.	of the same size generates the sampling distribution of the statistic which is being used to estimate the parent value.
5.2	Measures of reliability	Reliability refers to the consistency of a measure.	Reliability is how consistently a method measures some parameters. When one applies the same method to the same observation under the same conditions, one should get the same results.
6	**Timeliness**	**Statistics should be made available to users with the shortest delay possible and be delivered on the promised, advertised or announced dates. The timeliness of the data collection release to be compiled.**	
6.1	Timeliness	Length of time between data availability and the event or phenomenon they describe.	Timeliness refers to the speed of data availability, whether for dissemination or for further processing, and it is measured with respect to the time lag between the end of the reference period and the release of data. Timeliness is a crucial element of data quality: adequate timeliness corresponds to a situation where policy-makers can take informed decisions in time for achieving the targeted results. In quality assessment, timeliness is often associated with punctuality, which refers to the time lag between the release date of data and the target date announced in some official release calendar.
7	**Coherence and Comparability**	**Coherence is the degree to which data that are derived from different sources or methods, but refer to the same topic, are similar. Comparability is the degree to which data can be compared over time and domain**	

Item No	Concept name	Definition	Guidelines
7.1	Comparability – over time	Extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics.	Comparability aims at measuring the impact of differences in applied statistical concepts and definitions on the comparison of statistics between geographical areas, non-geographical dimensions, or over time. Comparability of statistics, i.e. their usefulness in drawing comparisons and contrast among different populations, is a complex concept, difficult to assess in precise or absolute terms. In general terms, it means that statistics for different populations can be legitimately aggregated, compared and interpreted in relation to each other or against some common standard. Metadata must convey such information that will help any interested party in evaluating comparability of the data, which is the result of a multitude of factors.
7.2	Coherence	Extent to which statistics are reconcilable with those obtained through other Data Sources or statistical domains.	This metadata element is used to describe the differences in the statistical results calculated on the basis of different statistical domains, or surveys based on different methodologies (e.g. between annual and short-term statistics or between social statistics and national accounts).
8	**Data Processing**	**Any processing undertaken to finalise the data**	
8.1	Source data type	Characteristics and components of the raw statistical data used for compiling statistical aggregates.	This metadata element is used to indicate whether the dataset is based on a survey, on administrative data sources, on a mix of multiple data sources or on data from other statistical activities. If sample surveys are used, some sample characteristics should also be given (e.g. population size, gross and net sample size, type of sampling design, reporting domain). If administrative registers are used, the description of registers should be given (e.g. source, primary purpose, etc.).
8.2	Frequency of data collection	Time interval at which the source data are collected.	The frequencies with which the source data are collected and produced could be different: a time series could be collected from the respondents at quarterly frequency but the data production may have a monthly frequency. The frequency of data collection should therefore be described.

Item No	Concept name	Definition	Guidelines
8.3	Mode and method of data collection method	The different combinations of data collection modes, pre-contact and follow-up modes. Method applied for gathering data for official statistics	The consolidation of an integrated management system underlies the feasibility of several management choices, such as those concerning mode changes for a given household. Three modes of data collection (Computer Assisted Web Interviewing (CAWI), Computer Assisted Telephonic Interviewing (CATI), and Computer Assisted Personal Interviewing CAPI) based on a common questionnaire (unimode design), and with both sequential and concurrent mode organization features. 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.
8.4	Data validation	Process of monitoring the results of data compilation and ensuring the quality of the statistical results.	Data validation describes methods and processes for assessing statistical data, and how the results of the assessments are monitored and made available to improve statistical processes. All the controls made in terms of quality of the data to be published or already published are included in the validation process. Validation also takes into account the results of studies and analysis of revisions and how they are used to improve statistical processes. In this process, two dimensions can be distinguished: (i) validation before publication of the figures and (ii) validation after publication.
8.5	Data compilation	Operations performed on data to derive new information according to a	In quality assurance frameworks, "Data compilation" refers to the description of statistical procedures used for producing intermediate data and final statistical outputs. Data compilation covers, among other things, the use of weighting schemes, methods for imputing missing values or

Item No	Concept name	Definition	Guidelines
		given set of rules.	source data, statistical adjustment, balancing/cross-checking techniques, and relevant characteristics of the specific methods applied.
8.6	Data identifier(s)	The unique identifier for an administered item within a registration authority.	Identifier is a sequence of characters, capable of uniquely identifying that with which it is associated, within a specified context. In order to exchange statistical information, a central institution has to agree with its partners about which statistical concepts are necessary for identifying the series (and for use as dimensions) and which statistical concepts are to be used as attributes. These definitions form the data structure definition. Each data structure definition has the following properties: a. identifier (providing a unique identification within an exchanged time series); b. name (also unique); and c. description (a description of the purpose and domain covered). The use of unique identifiers plays a pivotal role in the data architecture as it promotes interoperability by allowing individual records to be accurately linked across data sets (i.e., data linkage). Unique identifiers are alphanumeric codes or numbers assigned to individuals, cases or entities within a system. As implied by the term "unique", no two records should have the same identifier. When implementing unique identifiers within an institution, it is crucial to develop a standardized format with fixed naming conventions and implement robust validation checks to prevent the creation of invalid or duplicate identifiers. Clear guidance should also be in place regarding the management and retirement of unique identifiers.
9	**Metadata Update**	**The date on which the metadata element was inserted or modified in the database.**	

Item No	Concept name	Definition	Guidelines
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).

Sources relied upon for drawing definition and guidelines:

1. European Statistical System (ESS) Handbook for Quality and Metadata Reports — re-edition 2021)
2. Eurostat, "Technical Manual of the Single Integrated Metadata Structure (SIMS)", Luxembourg, 2014
3. Eurostat Statistics Explained, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Main_Page
4. FAO, https://www.fao.org/economic/the-statistics-division-ess/world-census-of-agriculture/conducting-of-agricultural-censuses-and-surveys/chapter-14-census-enumeration/en/
5. SDMX, "SDMX Glossary Version 2.1", December, 2020
6. Statistics Canada Quality Guidelines, "Defining Quality"