

Metadata Manual

Table of Contents

Introduction	3
Abstract	3
Metadata Users.....	4
Metadata.....	4
Disseminated statistical data.....	6
Managing Statistical Processes and Metadata.....	8
References.....	10

Introduction

Dubai Statistics Center "DSC" seeks to provide high-quality statistical data that fits users' needs in all economic and social fields in the Emirate of Dubai. Such data will definitely help to develop local development plans and contribute to providing the necessary data for the planner and user at the Federal and Gulf level. Additionally, they are featured to be comparable at the international level through accreditation and adaptation of international standards while preparing, analyzing and disseminating the official statistical data. Based on DSC's mission, vision and values and in accordance with the principle of transparency and meeting the users' needs and making them happy, DSC publishes the metadata on the statistics produced and disseminated in accordance with international standards in this field, as well as the basic principles of the official statistics issued by the United Nations which are linked to metadata including:

Principle 2 – Professional Standards and Professional Ethics

"To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage, and presentation of statistical data."

Principle 3 – Accountability and Transparency

"To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics."

Principle 5 – Sources of Official Statistics

"Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents."

Principle 9 – Use of International Standards

"The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels."

This manual intends to ensure the availability of metadata on statistical output at DSC to enhance transparency and support the quality of the statistical data.

Abstract

Metadata is about how, when, where, why, and by whom the data is collected. It is useful not only for statisticians who process, audit, and analyze data but also for data users. Data which is consistent, collected, and documented is essential for data users in order to understand outputs indications and statistical data quality. At the same, documenting concepts, definitions, terminologies, classifications, and methodologies used in collecting

data and describing the main aspects of data quality helps to effectively manage the statistics sources for statisticians and statistical agencies. Providing such data contributes and enhances transparency and increases trust at DSC and increases the number of users of DSC official website. Providing such data ensures the quality of statistical data and serves DSC users and employees. This will lead to enriching the work, providing resources and making the data and definitions easy to be accessed and reutilized.

Metadata Users

Metadata Users at DSC

This group includes many relative professions in stages of processing, producing, and disseminating official statistics and statistical data and functioning of the statistical information system; these professions are as follows: senior managers, information system designers, metadata content officers, IT specialist, dissemination specialists, planners, data auditors, and assessors.

End users at the national level

This group includes: government institutions, political decision makers, researchers, public employees, academics, librarians, journalists, enterprises, and the public.

Registered Data Providers

Their relationship with DSC entails providing metadata about data and information provided by them, so there is a need to provide metadata for information received by them in line with the method of processing such data.

International Users

Statistics international users such as multi-national companies, international organizations, and NGOs, etc. are significant users of statistical metadata. Using the international standards such as "metadata standard vocabularies" increases the probability of using statistical information by such users. Using international standards also helps in relieving burden on DSC when preparing reports to international organizations and IGOs.

Metadata

Defines and describes the disseminated data, in other words, such data is defined as essential data that helps users understand the nature of statistical data, values and tables disseminated which strengthen the optimal and correct usage of data.

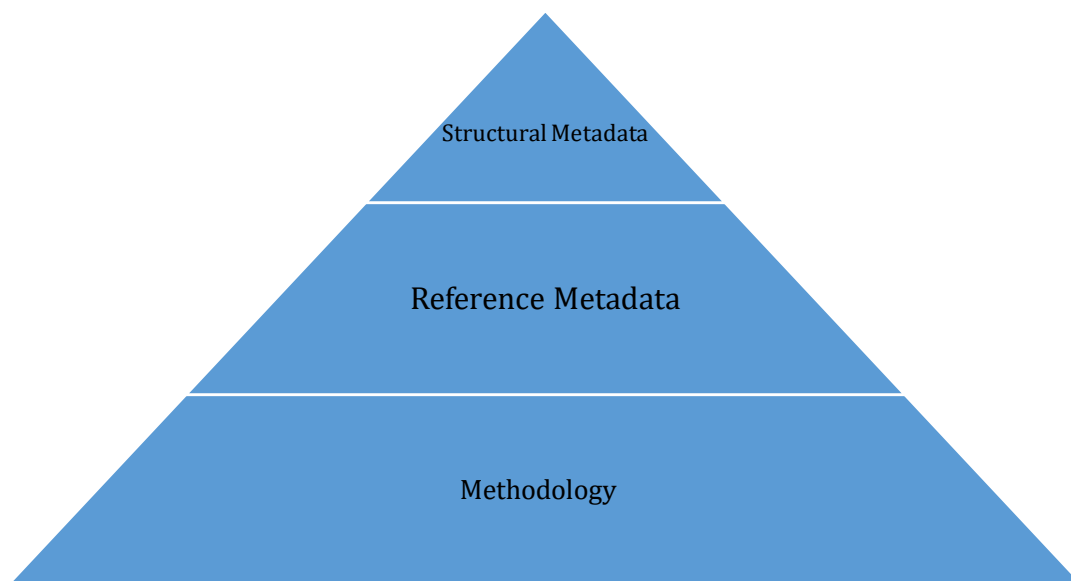
It is also the data which identifies and describes another data. It is defined as data about statistical data; it includes data and other documents which describe the official statistical outputs.

Types of Metadata

There are two types of metadata which are associated with statistical works and data:

1. Structural Metadata: It is the data that describes and defines data structure (such as the columns names in tables and its definition) in databases so that it is easy to be accessed and linked to each other; it is typically disseminated with tables and data.
2. Reference Metadata: It is data that is related to data contents. There are three types of reference metadata: conceptual metadata, this describes concepts being measured; systematic metadata, this describes the techniques used to generate data such as sampling techniques and collection; and finally quality of metadata, this describes data quality dimensions such as tameness and data accuracy.

Based on the international experiences in metadata field, the following hierarchical structure shows metadata hierarchy in statistics agencies:



Metadata becomes more detailed as we move from top to bottom and it has the following sequence:

First:

Structural Metadata: Comes at the top of the pyramid and it describes the statistical variables disseminated in the tables and it is usually about the primary variables, and it

provides the proper definition to understand the variable. There is no need for more details in this level.

Second:

Reference Metadata: Comes at the middle of the pyramid. It provides clarifications about, data, its source, frequency, and availability form. It also provides a brief description on how to process data and the methodological reference relied upon in processing data, and a description for data quality and reference. It must be disseminated and made available for all users, and updated whenever methodological need for update arises.

Third:

Methodology Metadata: Comes at the bottom of the pyramid. It provides a detailed description about the calculations used in producing statistical output including equations, weighting techniques, and different sources of data used in preparing statistical outputs and amendments to the approach and the measurement technique and any change to data processing and difference in the series. It clarifies the references used and relied upon in the methodology and concepts and classifications used and data collection methods and techniques used. It provides a detailed explanation for mechanisms and methodology of data accuracy confirmation and indicators of data quality and suitability. The methodology of each statistical output must be provided before the output is disseminated and be available for users with data or to be provided when needed.

Disseminated statistical data

Disseminated statistical data must have structural and reference metadata provided. Structural metadata considers the following aspects, wherever applicable:

1. Contact: Individual or organizational contact points for the data or metadata, including information on how to reach the contact points.
2. Metadata update: The date on which the metadata element was inserted or modified in the database.
3. Relevance: The degree to which statistical information meet current and potential needs of the users.
4. Statistical Indicator: Main characteristics of the indicator(its definition, classification, measurement method) in addition to specifying the units of measurements, the reference period, accuracy, and data source in which entity details is specified and the method of data collection.
5. Frequency and timeliness of dissemination: Describes the frequency of indicator dissemination and the length of time between data availability and the event or phenomenon they describe as well as the achieved dates for deliveries during a past period.
6. Coverage and comparability: Determines the geographical and time coverage of data in addition to the extent to which data is comparable geographically and over time.

7. Accessibility and clarity: The conditions and modalities by which users can obtain, use and interpret data.
8. Comment: Supplementary descriptive text, which can be attached to data or metadata.
9. Related metadata: any other related metadata.

Where reference metadata aspects include:

1. Contact: Individual or organizational contact points for the data or metadata, including information on how to reach the contact points.
2. Metadata update: The date on which the metadata element was inserted or modified in the database.
3. Statistical presentation: It covers main characteristics of the dataset, explained in a simple and understandable manner by referring to the published data and indicators, the used classifications, sector coverage, clarification of statistical concepts and definitions, the statistical unit that is covered and the collection of information about it, in addition to the study population, geographical coverage, coverage time and base period.
4. Unit of measure: The unit in which the data values are measured.
5. Reference period: The period of time or point in time to which the measured observation is intended to refer.
6. Institutional mandate: Set of rules or other formal set of instructions assigning responsibility as well as the authority to an organization for the collection, processing, and dissemination of statistics.
7. Confidentiality: A property of data indicating the extent to which their unauthorized disclosure could be prejudicial or harmful to the interest of the source or other relevant parties.
8. Release policy: Rules for disseminating statistical data to interested parties.
9. Frequency of dissemination: The time interval at which the statistics are disseminated over a given time period.
10. Accessibility and clarity: The conditions and modalities by which users can obtain, use and interpret data.
11. Quality management: Systems and frameworks in place within an organization to manage the quality of statistical products and processes.
12. Relevance: The degree to which statistical information meet current and potential needs of the users.
13. Accuracy and reliability: Closeness of computations or estimates to the unknown 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.
14. Timeliness and punctuality: describes the length of time between data availability and the event or phenomenon they describe, plus the time lag between the actual delivery of the data and the target date when it should have been delivered.
15. 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.

16. Cost and burden: Cost associated with the collection and production of a statistical product and burden on respondents.
17. Data revision: Any change in a value of a statistic released to the public.
18. Statistical processing: It includes the characteristics and components of raw statistical data used for compiling statistical aggregates, periodicity and collection method, in addition to data validation rules, and the procedures of values adjustment.
19. Comment: Supplementary descriptive text, which can be attached to data or metadata.

Managing Statistical Processes and Metadata

GSBPM v5.1 is used by DSC as a methodology for managing and documenting statistical processes which requires from all DSC's departments to document their main processes in line with processes stages and the requirements of each stage, and provide the relevant documents to facilitate producing and updating statistical projects metadata.

Needs	Design	Build	Data collection	Data processing	Analysis	Dissemination	Assessment
-------	--------	-------	-----------------	-----------------	----------	---------------	------------

- 1- **Identifying Needs:** This process includes identifying the need and feasibility of the statistical project and defining the needs and requirements through the feedback. This shall be made by identifying the main users and holding consultation with them; defining and documenting their needs; preparing and approving the project document proposal; preparing the project plan by identifying the plan activities and holding consultation with partners; identifying the needs of resources; and preparing and approving the budget.
- 2- **Design Stage:** This stage includes developing and designing the activities required to provide data so that all relevant techniques, methodologies, concepts, and data collection tools of statistical projects are reviewed. The stage also includes identifying objectives and specific objectives; analyzing available methodological options; defining specifications of the statistical project methodology; identifying the operational needs for collecting, processing, and disseminating data; analyzing the project tools of designing and testing the questionnaire and developing survey terms and definitions; and designing and approving the sample. This stage must be documented and provide such documents for purposes of quality control and providing metadata.
- 3- **Building Stage:** This stage includes the processes necessary to provide the tools and examine them to be ready for the actual implementation as the system is designed and programmed to implement the survey, data collection, testing and approval, and the approval of automated audit mechanisms within this stage.
- 4- **Data Collection:** This stage includes using the tools designed in the building stage, providing training and splitting researchers into work teams according to the

plan, conducting media campaign, collecting data, monitoring marking a response and or not and reasons for that, making sure of data accuracy during collection process in accordance with the established rules and finalizing data collection stage.

- 5- **Data Processing:** This stage includes processes and procedures of filtering data received from the field for purposes of processing and making it ready for the analysis stage. This includes auditing and coding questionnaires received from the field and entering their data into the system, verifying and auditing processes of entering the data, conducting the preliminary analysis of data and then comparing them with other sources.
- 6- **Analysis:** In this stage, the statistical data is analyzed, examined, compared and made sure of its safety and consistency with the technical standards in preparation for publication. This is done through multiple technical stages that include the overall audit of the results and their comparison with the previous surveys to ensure the logical, homogeneity and comparison with any other related surveys as well as with other available logistical sources available Also at this stage, the indicators that can be published are identified, the statistical tables are prepared, and a description of the main results is prepared.
- 7- **Dissemination:** This stage includes making results available to users after ensuring the quality of outputs according to the quality standards approved at DSC, and making sure that the dissemination standards are adopted, and reviewing and ensuring that the outputs report is correct in terms of language.
- 8- **Assessment:** This stage includes reviewing and assessing fieldwork procedures and their impact on response rates, assessing data processing, an assessing the methodologies used. The outputs of this stage are suggestions and recommendations for improving and updating work mechanisms.

Stages of managing statistical process mostly form a significant source for metadata of statistical projects and outputs. Therefore, documenting such stages significantly contribute to providing the metadata accurately and systematically in line with the dissemination time which entails considering the requirements of providing metadata during all statistical processes stages in a manner suits the concepts contained in this guide.

References

- Technologies for Data Exchange, Reference Metadata, UNITED NATIONS STATISTICAL DIVISION, 2015.
- Iraqi Metadata Toolbox, 2014
- Getting the Facts Right: A guide to presenting metadata with examples on Millennium Development Goal indicators, UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE, 2013.
- Framework of metadata requirements and roles in the S-DWH, Metadata Framework for Statistical Data Warehousing, Central Bureau for Statistics (CBS), 2013.
- The Role of Metadata in Statistics, U. S. Bureau of Labor Statistics, 2012.
- Understanding Metadata, National Information Standards Organization- NISO, 2012
- ESSnet on Data Warehousing, Statistics Sweden, 2011.
- Data and Metadata Reporting and Presentation Handbook, Organization for Economic Co-operation and Development- OECD, 2007.
- European Statistical System (ESS) Reference Metadata Reporting Standards. Retrieved August 21, 2019, from <https://ec.europa.eu/eurostat/data/metadata/metadata-structure>