AI-CATCH for Statistics

People Led, AI-Powered Change

Glossary of terms

Selected statistical terms

Prepared by the AI-CATCH Team
Mogens Grosen Nielsen and Lars Thygesen

Terms

Classification

a set of discrete, comprehensive and mutually exclusive observations that can be assigned to one or more variables which will be measured in the collation or presentation of the data. In the Statistical Data and Metadata eXchange (SDMX), “Classification systems” refers to a description of the classification systems being used and how they conform with internationally accepted standards guidelines, or good practices. It also refers to the description of deviations of classification systems compared to accepted statistical standards, guidelines, or good practices, when relevant. The terms “classification” and “nomenclature” are often used interchangeably, despite the definition of a “nomenclature” being narrower than that of a “classification”.

Continuity

refers both to the adequacy of resources and to the regulatory framework enabling the statistical operation to be conducted on a permanent basis.

Credibility

the trust users place in statistical products because they perceive that they are produced in a professional manner, in accordance with appropriate statistical standards, and that policies and practices are transparent.

Data and statistics

statistics are numerical information relating to an aggregate of data on units or observations. In general, this Guide uses the term “statistics” when referring to an output of a statistics production process and the term “data” when referring to an input or possibly throughput in the statistics production process (the term “data” includes microdata which, depending on the context, can be also an output).

Data ecosystem

a system in which a number of actors interact with each other to exchange, produce and utilize data. On a simple definition, a system can be understood as a set of connected parts forming a complex whole. There are multiple other definitions of a data ecosystem. The United Nations Development Programme model consists of data producers, data objects, infomediaries (i.e., media and other commercial information services) and data users, while other models put the national statistical office-led national statistical system at the centre of a system that consists of government agencies, academia and research institutions, the private sector, civil society and international and regional organizations.

Data providers and statistics producers

data providers, who provide an input to the statistics production process (such as respondents and holders or owners of statistical, administrative and other forms of data), and statistics producers, who produce a statistical output. Depending on the specific context, when using the term “data provider” this guide refers only to holders or owners of data.

Data sources

this guide distinguishes among three data sources according to their purpose and by the entity responsible for their compilation: (i) statistical data sources such as surveys, (ii) administrative data sources and
(iii) other data sources. In general, other data sources include data sources associated with the term “big data” unless already included, in some instances, in statistical or administrative data sources. New data sources can often be associated with other data sources; however, they may be considered part of statistical or administrative data sources as well, depending on national circumstances.

Efficiency

attainment of the maximum yield from a given level of resources in carrying out an activity.

Generic Activity Model for Statistical Organizations (GAMSO)

a model that extends and complements the Generic Statistical Business Process Model by modelling additional activities that support statistical production.

Generic Statistical Business Process Model (GSBPM)

a model that describes the processes used for the production of statistics, including the specification of needs, design, building, data collection, processing, analysis, dissemination of the products and evaluation of the process.

Interpretability

the ease with which users can understand, use and analyse data, considering their scope.

Metadata

data that define and describe other data. Structural metadata and reference metadata can be distinguished from each other. Structural metadata define and accompany the data and consist of identifiers and descriptors that are essential for discovering, organizing, retrieving and processing a statistical data set (e.g., titles, subtitles, short descriptions, dimension names, variable names, etc.). Reference metadata are of a more general nature and describe statistical concepts and methodologies used for the collection and generation of data and provide information on data quality, thereby assisting users with the interpretation of the data. Unlike structural metadata, reference metadata can be decoupled from the data (i.e., they can be generated, collected or disseminated separately from the statistics to which they refer).

National Quality Assurance Framework (NQAF)

a coherent and holistic system for statistical quality management that assures trust in and the quality of official statistics.

National statistical office

the leading statistical agency within a national statistical system. “National statistical office” and “national statistical institute” mean the same thing. In general, the national statistical office has a coordination role within the national statistical system and is responsible for the development, production and dissemination of official statistics across multiple statistical domains.

National statistical system (NSS)

the ensemble of statistical organizations and units (statistical agencies) within a country that develop, produce and disseminate official statistics on behalf of the national government (and other levels of government). It is the responsibility of each country to define the scope of its NSS (see also “statistical agencies”, “data providers and statistics producers” and “data ecosystem”).

Official statistics

statistics that describe, on a representative basis, economic, demographic, social and environmental phenomena of public interest. Official statistics are developed, produced and disseminated as a public good by the members of the NSS in compliance with the Fundamental Principles of Official Statistics and accepted quality frameworks such as the United Nations National Quality Assurance Framework, as well as other internationally agreed statistical

standards and recommendations. In many countries, official statistics are defined and described in the statistical programmes.

Open data

digital data that are made available with the technical and legal characteristics necessary for them to be freely used, reused and redistributed by anyone, at any time, anywhere. There are many similarities between the statistical quality principles of the United Nations National Quality Assurance Framework and the criteria for open data used in the Open Data Charter, such as timeliness and comprehensiveness, accessibility and usability, and comparability and interoperability.

Other statistics producers

entities that do not produce official statistics and are normally not members of the NSS. Other statistics producers have to be distinguished from other producers of official statistics, who are members of the NSS (see also “Statistical agencies”).

Principle, requirement, element to be assured

a principle is a general proposition, or procedure, to which statistical agencies and organizations are committed and that will guide them in meeting their quality-related objectives. A requirement is something needed in order to ensure the implementation of the United Nations National Quality Assurance Framework in chapter III. An element to be assured is a specific aspect of the National Quality Assurance Framework that identifies possible activities, methods and tools to meet the requirement. In this sense, an element to be assured reflects a good practice that is observed to work well in one or several national statistical offices or other producers of official statistics, and thus is a candidate to be promoted for use in other statistical agencies.

Quality

the degree to which a set of inherent characteristics of an object fulfils requirements. A simple definition of quality is “fit for use” or “fit for purpose”. It is the users’ needs that define the quality. Different users may have different needs that must be balanced against each other.

Quality assessment

the part of quality assurance that focuses on an assessment of how well quality requirements (the stated needs or expectations) are fulfilled.

Quality assurance

a planned and systematic pattern of all the actions needed to provide adequate confidence that a process and a product conform to stated requirements.

Quality management

the set of systems and frameworks in place within an organization to manage the quality of statistical products and processes. In the case of a national statistical office and other producers of official statistics, quality management also includes managing the statistical system and the institutional environment, as applicable. Quality management includes quality assurance, but the terms are often used interchangeably; quality management is a more overarching concept, while quality assurance implies a greater focus on concrete actions.

Quality policy

a document that defines top management’s commitment to quality. A quality policy statement should describe the overall quality orientation of an organization and clarify its basic intentions. Quality policies should be used to generate quality objectives and serve as a general framework for action. Quality policies can be based on the ISO 9000 Quality Management Principles and should be consistent with the organization’s other policies.

Respondent burden

the effort, in terms of time and cost, required for respondents to provide satisfactory answers to a survey.

Respondents

enterprises, authorities, individuals and others from whom data and associated information are collected for use in the production of statistics.

Revision

a change in a value of statistics released to the public. Changes can be the result of errors, but normally the term “revision” is reserved for planned changes in published numbers. Statistics can be revised when more and better source data become available, or because of a change in methodology.

Risk management

the identification, analysis, assessment, control and avoidance, minimization or elimination of unacceptable events.

Source data

data collected (from respondents, administrative entities and other data providers) by members of the national statistical system to be used in the compilation and production of official statistics.

Standard

a normative document, established by a consensus and approved by a recognized body, that specifies rules, guidelines or characteristics for common and repeated use in relation to activities or their results with the aim of achieving the optimum degree of order in a given context.

Statistical agencies

the members of the national statistical system (NSS) are known as “statistical agencies” and include the national statistical office and other producers of official statistics. In general, statistical agencies other than the national statistical office are geared primarily towards purposes and tasks other than the production of official statistics, and only a section or a small group of persons within the institution produces statistics.

The process and output quality requirements are the same for all official statistics. In the case of a ministry or administrative agency that is not exclusively tasked with the production of statistics as a mission activity, however, the requirements associated with the institutional environment only apply to the section engaged in the production of official statistics. Thus, while the ministry or administrative agency is generally not independent, the unit within the administrative agency responsible for producing statistics must decide how to produce and when to disseminate its statistics independently.

Statistical Data and Metadata eXchange (SDMX)

an international initiative that aims at standardizing and modernizing ("industrializing") the mechanisms and processes for the exchange of statistical data and metadata among international organizations and their member countries.

Statistical purpose

tasks aimed at developing, producing and disseminating official statistics, including experimenting and testing.

Statistical standards

standards define and establish uniform specifications and characteristics for products or services. In the context of this Guide "statistical standards" refers to a comprehensive set of statistical concepts, definitions, classifications and models, methods and procedures used to achieve the uniform treatment of statistical issues within or across processes and across time and space.

Transparency

the information context in which data are provided to users, together with metadata (explanations, documentation, quality information that may limit the use of the data).