Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

10-Dec-2012

English - Or. English

Statistics Directorate
Committee on Statistics

STD/CSTAT/MICRO(2012)5
Unclassified

Expert Group for International Collaboration on Microdata Access

GLOSSARY OF TERMS - MICRODATA ACCESS
DRAFT COLLECTION OF RELEVANT TERMS

13-14 December 2012
OECD Conference Centre
Paris, France

This draft was prepared by the Secretariat.

OECD Contact: Young-tae Son, email: young-tae.son@oecd.org

JT03332530

English - Or. English

GLOSSARY OF TERMS - MICRODATA ACCESS. DRATF COLLECTION OF RELEVANT TERMS

Introductory note

1. At its first meeting in June 2012, the "OECD Expert Group for International Collaboration on Microdata Access" recognised the need of clearly establishing the terminology to be used in relation to the access to and exchange of microdata. In fact, since the first sessions of the June meeting it appeared evident that there was no common understanding of terms, not even of basic concepts/terms that are usually utilised in the context of microdata; an illustrative example is the term "remote access", which experts from different countries and statistical agencies used, at the meeting, with heterogeneous meanings. Interestingly, although an international definition of the term "microdata" exists, some experts interpreted it as a narrower or, on the opposite, broader concept. In light of this and noticing that no such standardised glossary of terminology on microdata currently exists at the international level, the Expert Group decided to compile a "Glossary of Terms" to submit to CSTAT as the international reference terminology in the area of microdata.

The purpose of the draft collection

2. The purpose of this draft document, which was prepared by the Secretariat, is to provide an initial collection of terms to serve as a basis for a comprehensive Glossary of relevant terms. The draft is submitted to the careful examination of the Expert Group. Experts are invited to accurately review each "entry" in the document and agree on:

- *The final selection of terms to be retained*. This might imply adding entries not yet included in the draft collection, and deleting entries that are deemed not to be central to the area of microdata.

- *The precise definition of each term retained*. This might involve for some terms a revision of existing definitions, or the setting up of entirely new definitions.

Criteria for identifying relevant terminology

3. The draft collection covers all the terms that the Secretariat identified as part of a key vocabulary on microdata. Specifically, the criteria to retain "terms" were the following:

a. Terms used in official methodological and terminology documents of national statistical offices (NSOs) and international organisations;

b. Terms used in the previous meeting of the OECD Expert Group;

c. Additional basic terms that seem to pertain to the area of microdata.

Presentation of the selected terms

4. In the document, the terms are presented according to the following criteria:

a. By alphabetical order;

b. With definition, context, source(s) and hyperlink(s), if applicable;

c. When a term is reported by several sources, all the different definitions of the same term are reproduced;

d. Some terms with no definition are included, if they appear to be part of a commonly used vocabulary in the field of microdata.

5. Finally, the text contains cross-references when two different terms refer to the same concept.

Sources

6. For the preparation of this draft, only documents from NSOs and international organisations' sources were consulted. In particular, the terms included in this draft collection were drawn from five main references:

- International Household Survey Network - IHSN (2010), *Dissemination of Microdata Files - Principles, Procedures and Practices*, Dupriez O. and E. Boyko, IHSN Working Paper No. 005. http://www.ihsn.org/home/index.php?q=focus/dissemination-microdata-files-principles-procedures-and-practices

- OECD (2007), *Glossary of Statistical Terms*, http://stats.oecd.org/glossary/index.htm

- OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding,* http://www.oecd.org/science/scienceandtechnologypolicy/38500813.pdf

- UNECE (2007), *Managing Statistical Confidentiality & Microdata Access – Principles ad Guidelines of Good Practice*, http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

- UNECE (2000), Conference of European Statisticians Statistical Standards and Studies – No. 53, "*Terminology On Statistical Metadata*" http://www.unece.org/stats/publications/53metadaterminology.pdf

7. To be noted, the original sources of the definitions of the terms drawn from the *OECD Glossary of Statistical Terms* include:

- Eurostat (1996), "*Manual on disclosure control methods*", Office for Official Publications of the European Communities, Luxembourg, http://ec.europa.eu/eurostat/ramon/statmanuals/files/manual_on_disclosure_control_methods_1996.pdf

- IMF (2000), *Code of Good Practices on Transparency in Monetary and Financial Policies*, Part 1-Introduction; approved by the IMF Executive Board on July 24, http://www.imf.org/external/np/mae/mft/sup/part1.htm#appendix_III

- IMF, "Guide to the Data Dissemination Standards, Module 1: The Special Data Dissemination Standard", Washington, May 1996, http://dsbb.imf.org/sddsindex.htm

- ISO/IEC FDIS 11179-1, "Information technology - Metadata registries - Part 1: Framework", March, http://metadata-standards.org/11179/

- OECD, IMF, ILO - Interstate Statistical Committee of the Commonwealth of Independent States (2002), *Measuring the Non-Observed Economy: A Handbook*, Annex 2, Glossary, http://www.oecd.org/std/nationalaccounts/1963116.pdf .

- Statistics Canada (1998), *Statistics Canada Quality Guidelines*, 3rd edition, October, http://www.statcan.ca/english/freepub/12-539-XIE/12-539-XIE.pdf

- Statistical Data and Metadata Exchange (SDMX) – BIS, ECB, Eurostat, IBRD, IMF, OECD and UNSD – Metadata Common Vocabulary, http://www.sdmx.org/

- Statistics New Zealand, "Classifications and Standards", http://www.stats.govt.nz/surveys_and_methods/methods/classifications-and-standards.aspx

- UNECE (1995), *Guidelines for the Modelling of Statistical Data and Metadata*, Conference of European Statisticians, Methodological material, United Nations, Geneva, http://www.unece.org/stats/publications/metadatamodeling.pdf

- UNESCO, OECD, Eurostat (2002), *Data Collection on Education Systems: Definitions, Explanations and Instructions*, http://circa.europa.eu/Public/irc/dsis/edtcs/library?l=/public/unesco_collection/2002/uoe02def_pdf/_EN_1.0_&a=d

- UNSC and UNECE (1995), Guidelines for the Modelling of Statistical Data and Metadata, Conference of European Statisticians Methodological Material, page 1, http://www.unece.org/fileadmin/DAM/stats/publications/metadatamodeling.pdf

- United States Bureau of the Census, Software and Standards Management Branch, Systems Support Division, *Survey Design and Statistical Methodology Metadata*, Washington D.C., August 1998, Section 3.3.17, http://www.census.gov/srd/www/metadata/metada18.pdf

- Dekker A., "Adapting new technologies to census operations", Symposium on Global Review of 2000 Round of Population and Housing Censuses: Mid-Decade Assessment and Future Prospects, Statistics Division, Department of Economic and Social Affairs, United Nations Secretariat New York, 7-10 August 2001, Glossary, http://unstats.un.org/unsd/demographic/docs/symposium_06.htm

COLLECTION OF TERMS

Academic staff

Academic Staff (International Standard Classification of Education (ISCED) 5-6) includes personnel whose primary assignment is instruction, research, or public service. This includes staff personnel who hold an academic rank with titles such as professor, associate professor, assistant professor, instructor, lecturer, or the equivalent of any of these academic ranks. The category includes personnel with other titles, (e.g. dean, director, associate dean, assistant dean, chair or head of department), if their principal activity is instruction or research. It does not include student teachers or teacher aides

Source: 2001 Data Collection on Education Systems: Definitions, Explanations and Instructions, UNESCO, OECD, Eurostat, page 45.

Access arrangements

See **Formal responsibility**

Accessibility (as a statistical data quality dimension)

The ease and the conditions with which statistical information can be obtained.

Context: Accessibility refers to the availability of statistical information to the user (International Monetary Fund, "Data Quality Assessment Framework - DQAF - Glossary").

Accessibility includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of accessibility for some users. (Statistics Canada, "Statistics Canada Quality Guidelines", 4th edition)

In SDMX, "Accessibility of Documentation" refers to the availability of documentation of various aspects of the data (sources and methods documents) and the content of such documentation.

Source: Statistical Data and Metadata Exchange (SDMX) – BIS, ECB, Eurostat, IBRD, IMF, OECD and UNSD – Metadata Common Vocabulary
http://www.sdmx.org/

Accountability

The performance of data access arrangements should be subject to periodic evaluation by user groups, responsible institutions and research funding agencies. Although each party is likely to use somewhat different evaluation criteria, the sum total of the results should provide a comprehensive picture of the value of data and of data access regimes. Such evaluations should help to increase the support for open access among the scientific community and society at large.

The following should be considered in establishing evaluation criteria:
• Overall public investments in the production and management of research data.
• Management performance of data collection and archival agencies.
• Extent of re-use of existing data sets.
• Knowledge generated from the re-use of existing data.
• The use of targeted foresight exercises to determine the nature and scope of data preservation activities and the types of data most likely to be needed in the future. Even if gaining clear insight into the cost, benefit and performance of data access arrangements will not be an easy task, those in charge of data access arrangements should put effort into showing the benefits of open data access to justify and help ensure sustained support from all levels of government.

Source: OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, Paris, 2007, p. 21

Administrative data

Administrative data is the set of units and data derived from an administrative source.

Source: OECD, IMF, ILO, Interstate Statistical Committee of the Commonwealth of Independent States, "Measuring the Non-Observed Economy: A Handbook", Annex 2, Glossary, Paris, 2002
Hyperlink: http://www.oecd.org/dataoecd/9/20/1963116.pdf

Administrative data / Administrative records

The data collected by sources external to statistical offices

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Administrative source

Administrative source is the organisational unit responsible for implementing an administrative regulation (or group of regulations), for which the corresponding register of units and the transactions are viewed as a source of statistical data.

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
Hyperlink: http://www.oecd.org/dataoecd/9/20/1963116.pdf

Anonymised microdata files - licensed files

The term anonymised implies that not only are names and addresses removed but that other steps are taken to ensure that identification of individuals is highly unlikely. Licensed files are distinct from Public Use Files in that use is restricted to approved researchers for approved purposes. A legal undertaking is signed before files are provided to them.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 106
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Anonymised microdata files - Public Use Files

These are microdata files that are disseminated for general public use. They have been anonymised and are often released on a medium such as CD-ROM sometimes through a data archive. The term anonymised implies that not only are names and addresses removed but that other steps are taken to ensure that identification of individuals is highly unlikely.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 106
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Attribute

A characteristic of an object or entity.

Context: An entity is any concrete or abstract thing of interest, including associations among things. A composite attribute is an attribute whose data type is non-atomic. An attribute instance is a specific instance of an attribute. An attribute value is the value associated with an attribute instance (ISO/IEC International Standard 11179-3 "Information technology - Metadata registries (MDR)-Part 3: Registry metamodel and basic attributes", February 2003). Within SDMX, a data or metadata attribute is a statistical concept providing qualitative information about a specific statistical object such as a data set, observation, data provider, or dataflow.

Concepts such as units, magnitude, currency of denomination, titles and methodological comments can be used as attributes in the context of an agreed data exchange. A conditional attribute is permitted to take empty values. A mandatory attribute is an attribute which must take a value, otherwise the corresponding observation, which it refers to, is not considered as meaningful enough, e.g. with regard to the "status" of an observation or the units in which a whole time series is expressed. Within the SDMX information model, attribute value is the value of an attribute, such as the instance of a coded or uncoded attribute in the context of a data structure.

Source: ISO/IEC FDIS 11179-1 "Information technology - Metadata registries - Part 1: Framework", March 2004

Attribute

An inherent characteristic of an object

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Confidential cells

The cells of a table which are non-publishable due to the risk of statistical disclosure are referred to as confidential cells.

Context: By definition there are three types of confidential data where disclosure might happen and therefore cells that are confidential:

Small counts. A tabular cell is confidential, if less than m entities contribute to the total of that cell. The value of m is called a threshold and is usually determined by the statistical authority according to the desired degree of confidentiality protection: m is at least 3 but sometimes m= 5 is given. In the case of a threshold of m=3, a cell is confidential if the figure in the cell shows the data of only one unit, or if the figure is the sum of two entities and one respondent has the possibility of disclosing the figure of the other respondent by subtraction of his own figure from the sum. This is also known as the threshold rule.

Dominance or case of predominance. (a) dominance rule, concentration rule, (n,k) rule: A cell is regarded as confidential, if the n largest units contribute more than k% to the cell total. The n and k are given by the statistical authority and differ quite a lot, *e.g.* you find that n=2 and k=85, which means that a cell is defined as confidential if the two largest units contribute more than 85% to the cell total. (b) prior posterior ambiguity rule, p/q rule: it is assumed that out of publicly available information the contribution of one individual to the cell total can be estimated to within p per cent (p=error before publication); after the publication of the statistic the value can be estimated to within q percent (q=error after publication). In the p/q rule the ratio p/q represents the information gain through publication and in the prior posterior ambiguity rule the difference p-q. If the information gain is unacceptable the cell is declared as confidential. P and q are given by the statistical authority and thus the definition of the acceptable level of information gain.

Secondary confidentiality/derivation: Even if all confidential cells containing small counts or cases of predominance are protected by disclosure control methods (=primary protection), disclosure might be possible by recalculating confidential cells as the difference between a total and the sum of cells corresponding to that total. This recalculation of primary protected cells is called derivation. Derivation can occur (a) within one two-dimensional table or higher-dimensional tables, when margin totals are given in the lines, the columns or in a set of lines or columns; (b) between tables and subtables in the case of three or more dimensions e.g. between geographic levels or between aggregation levels (total economy, sector); (c) between different tables on the same aggregation or geographical level containing different sorts of information.

Note: Small counts and dominance are collectively primary confidentiality.

Source: Eurostat, "Manual on disclosure control methods", Office for Official Publications of the European Communities, Luxembourg, 1996, p. 8-9

Confidential data

Confidential data are data which are subject to confidentiality clauses.

Context: The data collected by many national statistical agencies are subject to national rules regarding confidentiality. The two main reasons for declaring data to be primary confidential are: a) too few units in a cell; b) dominance of one or two units in a cell. The limits of what constitutes "too few" or "dominance" vary between statistical domains. In the European Union, confidential data is defined in Article 13 of Council Regulation No 322/97, as:

1. Data used by the national authorities and the Community authority for the production of Community statistics shall be considered confidential when they allow statistical units to be identified, either directly or indirectly, thereby disclosing individual information. To determine whether a statistical unit is identifiable, account shall be taken of all the means that might reasonably be used by a third party to identify the said statistical unit.

2. By derogation from paragraph 1, data taken from sources which are available to the public and remain available to the public at the national authorities according to national legislation, shall not be considered confidential.

Source: Statistical Data and Metadata Exchange (SDMX) - BIS, ECB, Eurostat, IBRD, IMF, OECD and UNSD - Metadata Common Vocabulary
*Hyperlink:*www.sdmx.org

Confidentiality

Data confidentiality is a property of data, usually resulting from legislative measures, which prevents it from unauthorized disclosure.

Context: In SDMX, "Confidentiality" refers to the legislative measures or other formal provision which prevent unauthorised disclosure of data that identify a moral or physical person either directly or indirectly. Also refers to the procedures in place to prevent disclosure of confidential data, including rules applying to staff, aggregation rules when disseminating data, provision of unit records, etc.

Source: Economic Commission for Europe of the United Nations (UNECE), "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000
Hyperlink: http://www.unece.org/stats/publications/53metadaterminology.pdf

Data

The physical representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means.

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Database

A data file or set of data with relationships expressed among data. Data stored in the database are independent of any particular application

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Data confidentiality

A property of data, usually resulting from legislative measures, which prevents it from unauthorized disclosure

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Data Cubes -Netherlands

Data cubes are the main vehicle for releasing all statistical information. Statistical confidentiality protection is applied in a routine fashion. Moreover, data cubes can be easily linked and compared on a meso level. Conversely, a lack of coherence is easily discovered. Adding data cubes to the StatLine database ensures that statistical information is produced and published to serve the public at large. Data cubes are primarily made and used to serve the public at large. Even if they are produced and paid for by a third party, as a matter of policy the resulting data cubes are available for all.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 32
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Data dissemination

Dissemination is the release to users of information obtained through a statistical activity.

Context: Data dissemination consists of distributing or transmitting statistical data to users. Various release media are possible; for example: electronic format including the internet, CD-ROM, paper publications, files available to authorised users or for public use; fax response to a special request, public speeches, press releases. Under the SDDS, the concept of dissemination formats is divided into two categories: "hardcopy" and "electronic" publications, which detail the reference documents through which users may access the data described in the metadata and, where relevant, detailed components beyond the minimum prescribed.

In SDMX, "Supplementary Data" refers to a description of data not routinely disseminated that are made available to users upon request. It may include customized tabulations that can be provided (perhaps for a fee) to meet specific requests. Also include information on procedures for obtaining these supplementary data.

Source: Statistics Canada, "Statistics Canada Quality Guidelines", 3rd edition, October 1998, page 59
Hyperlink: http://www.statcan.ca/english/freepub/12-539-XIE/12-539-XIE.pdf

Data Enclave

This is a facility equipped with computers not linked to the internet or an external network and from which no information can be downloaded via USB ports, CD-DVD or other drives. Data enclaves contain data that are particularly sensitive or allow direct or easy identification of respondents. Examples include complete population census datasets, enterprise surveys and certain healthrelated datasets containing highly-confidential information. Users interested in accessing a data enclave will not necessarily have access to the full dataset – only to the particular data subset they require.

Source: Olivier Dupriez and Ernie Boyko. 2010, “Dissemination of Microdata Files; Principles, Procedures and Practices” IHSN[1] Working Paper No. 005, August 2010, p. 7
http://www.ihsn.org/home/index.php?q=focus/dissemination-_microdata-files-principles-__procedures-and-practices

[1] International Household survey Network

Data file

An organized collection of related records of data

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Data laboratories

This involves working on-site at the National Statistical Office, or one of its Branches, to obtain access to microdata. Access could be direct or indirect through staff of the National Statistical Offices. If access is direct, the researcher is in effect being treated as a temporary employee of the National Statistical Office with the inherent responsibilities.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 107
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Data processing

The operation performed on data in order to derive new information according to a given set of rules

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Data protection

An activity aimed at covering or shielding data from physical damage or unauthorized access

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Data security

The measures taken to prevent unauthorized access or use of data

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Data set

Any organized collection of data

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Data type

A category used to classify the collection of letters, digits, and/or symbols to depict values of a data element based upon the operations that may be performed on the data element.

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Disclosure analysis

Disclosure analysis is the process of protecting the confidentiality of data. It involves limiting the amount of detailed information disseminated and/or masking data via noise addition, data swapping, generation of simulated or synthetic data, etc.

Source: United States Bureau of the Census, Software and Standards Management Branch, Systems Support Division, "Survey Design and Statistical Methodology Metadata", Washington D.C., August 1998, Section 3.3.17, page 28.
Hyperlink: http://www.census.gov/srd/www/metadata/metada18.pdf

Disclosure control

The complex of measures preventing unauthorized access to sensitive statistical information.

Source: Adapting new technologies to census operations, A. Dekker, Symposium on Global Review of 2000 Round of Population and Housing Censuses: Mid-Decade Assessment and Future Prospects, Statistics Division, Department of Economic and Social Affairs, United Nations Secretariat New York, 7-10 August 2001, Glossary
Hyperlink: http://unstats.un.org/unsd/demographic/docs/symposium_06.htm

Efficiency

One of the central goals of promoting data access and sharing is to improve the overall efficiency of publicly funded scientific research to avoid the expensive and unnecessary duplication of data collection efforts. Consideration should be given to the following:

• Data access arrangements should promote further cost effectiveness within the global science system by describing good practices in data management and specialised support services.

• While publicly funded research data are subject to the default rule of openness under “Principle A. Openess”, this does not mean that all such data should be preserved permanently.The data archiving community should carry out cost-benefit assessments periodically and constantly develop and refine retention protocols to ensure that those data sets with the greatest potential utility are preserved and made accessible. Use of accepted retention protocols and thorough documentation of data should help to reduce unnecessary duplication of effort as well as to establish the necessary selectivity in preservation.

• Specialised support services, for example through collaboration with non-academic specialists on specific research projects or the engagement of data management specialist organisations, should be considered as a means to ensure the cost-effective production, use, management and archiving of research data.

• Insufficient incentives for researchers or database producers may lessen their efforts on data-related activities. The development of new reward structures and the adaptation of existing ones, including recognition of data management activities in tenure and promotion review, should be considered as a way to address this problem.

Source: OECD (2007), "*OECD Principles and Guidelines for Access to Research Data from Public Funding*", Paris, p. 21.

Formal responsibility - Access arrangements

Access arrangements should promote explicit, formal institutional practices, such as the development of rules and regulations, regarding the responsibilities of the various parties involved in data-related activities. These practices should pertain to authorship, producer credits, ownership, dissemination, usage restrictions, financial arrangements, ethical rules, licensing terms, liability, and sustainable archiving.

Access arrangements, whether at the governmental or institutional levels, should be developed in consultation with representatives of all directly affected parties. In collaborative research programmes or projects, and especially in international scientific co-operation or in research projects based on public/private partnerships where there are differences in regulatory frameworks, the parties involved should negotiate research data sharing arrangements as early as possible in the life of the research project, ideally at the initial proposal stage. This will help ensure that adequate and timely consideration will be given to issues such as the allocation of resources for sharing and sustainable preservation of research data, differences in national intellectual property laws, limitations due to national security, and the protection of privacy and confidentiality.

Access arrangements also should be responsive to factors such as the characteristics of the data, their potential value for research purposes, the level of data processing (raw versus partially processed versus final), whether they are homogeneous data from a facility instrument or sensor versus heterogeneous field data collected by single researchers, data on human subjects or physical parameters, and whether the data are generated directly by a government entity or as a result of government funding. These variations in the origin or type of data should be taken into consideration when establishing data access arrangements. Further, consideration should be given to the following:

- Many of the problems related to access, dissemination and sharing of data result from the lack of explicit institutional agreements on the terms of access and use. With data management becoming ever more complex in certain areas of research, traditional informal arrangements between researchers may no longer be adequate and may need to be complemented by formally agreed practices and procedures.
- Responsibility for the various aspects of data access and management should be established in relevant documents, such as descriptions of the formal tasks of institutions, grant applications, research contracts, publication agreements, and licenses.
- Long-term sustainability of the infrastructure required for data access is particularly important. Research institutions and government organisations should take formal responsibility for ensuring that research data are effectively preserved, managed and made accessible in order that they can be put to efficient and appropriate use over the long term.

Source: OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, Paris, p. 17.

Internal access

Internal access refers to giving full transparency to any necessary pre-release access within government, as deemed appropriate by the government.

Context: Under the SDDS, this entails the listing of persons or officials holding designated positions within the government, but outside the agency producing the data, who have pre-release access to the data and the reporting of the schedule according to which they receive access.

Source: Statistical Data and Metadata Exchange (SDMX) - BIS, ECB, Eurostat, IBRD, IMF, OECD and UNSD - Metadata Common Vocabulary
Hyperlink: www.sdmx.org

International statistical standard

The term international statistical standard refers to the comprehensive body of international statistical guidelines and recommendations that have been developed by international organisations working with national agencies.

Context: The formulation of international statistical standards necessarily entails an extensive process of consultation and discussion between international organisations and between international organisations and their member countries.

The standards cover almost every field of statistical endeavour from data collection, processing and dissemination and almost very statistical subject. Such standards also include international statistical classifications. The most comprehensive database of existing international statistical guidelines and recommendations is maintained on the United Nations Statistical Division website, the Methodological publications in statistics. This database also lists standards currently being developed by international organisations.

Source: Statistical Data and Metadata Exchange (SDMX) - BIS, ECB, Eurostat, IBRD, IMF, OECD and UNSD - Metadata Common Vocabulary
Hyperlink: www.sdmx.org

Interoperability

Technological and semantic interoperability is a key consideration in enabling and promoting international and interdisciplinary access to and use of research data. Access arrangements, should pay due attention to the relevant international data documentation standards. member countries and research institutions should co-operate with international organizations charged with developing new standards. Although science is becoming a highly globalised endeavour, incompatibility of technical and procedural standards can be the most serious barrier to multiple uses of data sets. Factors that should be considered include:

- The standards employed should be explicitly mentioned as this is the first requirement for interoperability.
- Adoption of the practices of disciplines most advanced in this respect should be promoted, in particular by the international professional organisations dealing with science and the collection and preservation of data for research and technological purposes.
- The work of organisations engaged in setting more general information and communication technology standards should also be considered.

Source: OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, Paris, p. 19.

Key variables

Key variables allow users to link information from one file to that of another file.

Source: Olivier Dupriez and Ernie Boyko. 2010, "Dissemination of Microdata Files; Principles, Procedures and Practices" IHSN Working Paper No. 005, August 2010, p. 3

http://www.ihsn.org/home/index.php?q=focus/dissemination- microdata-files-principles- procedures-and-practices

Keyword

Word used for linking to certain classified objects.

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Legal conformity

Data access arrangements should respect the legal rights and legitimate interests of all stakeholders in the public research enterprise. Access to, and use of, certain research data will necessarily be limited by various types of legal requirements, which may include restrictions for reasons of:

• National security: data pertaining to intelligence, military activities, or political decision making may be classified and therefore subject to restricted access.

• Privacy and confidentiality: data on human subjects and other personal data are subject to restricted access under national laws and policies to protect confidentiality and privacy. However, anonymisation or confidentiality procedures that ensure a satisfactory level of confidentiality should be considered by custodians of such data to preserve as much data utility as possible for researchers.

• Trade secrets and intellectual property rights: data on, or from, businesses or other parties that contain confidential information may not be accessible for research.

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

• Legal process: data under consideration in legal actions (sub judice) may not be accessible.

Subscribing to professional codes of conduct may facilitate meeting legal requirements.

Source: OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, Paris, p. 16

Licensed Files

Licensed Files – also called *Research Files* –are distinct from PUFs: their dissemination is restricted to users who have received authorization to access them after submitting a documented application and signing an agreement governing the data's use. While typically licensed files are also anonymised to ensure the risk of identifying individuals is minimised when used in isolation, they may contain potentially dentifiable data if linked with other data files.3 Direct identifiers such as respondents' names must be removed from a licensed dataset. The data files may, however, still contain indirect variables that could identify respondents by matching them to other data files such as voter lists, land registers or school records.

Source: Olivier Dupriez and Ernie Boyko. 2010, "Dissemination of Microdata Files; Principles, Procedures and Practices" IHSN Working Paper No. 005, August 2010, p. 7
http://www.ihsn.org/home/index.php?q=focus/dissemination- microdata-files-principles- procedures-and-practices

Macrodata

See **Tabular data**

Macrodatabase

The database containing macrodata

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000

Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Microdata - 1

The term microdata can refer to data about an individual person, household, business or other entity. It may be data directly collected by the NSO or obtained from other sources, such as administrative sources.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 1
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Microdata - 2

Data at the level of the individual respondent.

Source: Olivier Dupriez and Ernie Boyko. 2010, "Dissemination of Microdata Files; Principles, Procedures and Practices" IHSN Working Paper No. 005, August 2010, p. 1
http://www.ihsn.org/home/index.php?q=focus/dissemination-_microdata-files-principles-_procedures-and-practices

Microdatabase

The database containing microdata

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

National Statistical Institute (NSI)

Same as National Statistical Office – 1 ?

National Statistical Office (NSO) -1

The national statistical office is the leading statistical agency within a national statistical system.

Source: Measuring the Non-Observed Economy: A Handbook, OECD, IMF, ILO, Interstate Statistical Committee of the Commonwealth of Independent States, 2002, Annex 2, Glossary.
Hyperlink: http://www.oecd.org/dataoecd/9/20/1963116.pdf

National Statistical Office (NSO) -2

Although the term is used in the singular, it is meant to incorporate all statistical agencies, or statistical units within government departments, who produce official statistics and provide access to microdata for statistical or research purposes.

Source: UNECE, "*Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice*", 2007, p. 106
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

National statistical system (NSS)

The national statistical system (NSS) is the ensemble of statistical organisations and units within a country that jointly collect, process and disseminate official statistics on behalf of national government.

Source: Measuring the Non-Observed Economy: A Handbook, OECD, IMF, ILO, Interstate Statistical Committee of the Commonwealth of Independent States, 2002, Annex 2, Glossary.
Hyperlink: http://www.oecd.org/dataoecd/9/20/1963116.pdf

Official statistics

Official statistics are statistics disseminated by the national statistical system, excepting those that are explicitly stated not to be official.

Source: Measuring the Non-Observed Economy: A Handbook, OECD, IMF, ILO, Interstate Statistical Committee of the Commonwealth of Independent States, 2002, Annex 2, Glossary.
Hyperlink: http://www.oecd.org/dataoecd/9/20/1963116.pdf

On-line access

See Remote access facilities

On-site access

Off-line access

Off-site access

Openness

Openness means access on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination. Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.

Source: OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding* , Paris, p. 15

Passive metainformation system

A passive metainformation system contains only references to data, not the data themselves.

Source: Guidelines for the Modelling of Statistical Data and Metadata, UNECE and UNSC, Conference of European Statisticians Methodological Material, 1995 – page 1.

Perturbation-based methods

Perturbation-based methods falsify the data before publication by introducing an element of error purposely for confidentiality reasons. This error can be inserted in the cell values after the table is created, which means the error is introduced to the output of the data and will therefore be referred to as output perturbation, or the error can be inserted in the original data on the microdata level, which is the input of the tables one wants to create; the method with then be referred to as data perturbation - input perturbation being the better but uncommonly used expression. Possible methods are:

rounding;
random perturbation;
disclosure control methods for microstatistics applied to macrostatistics.

Source: Eurostat, 1996, "Manual on disclosure control methods", Office for Official Publications of the European Communities, Luxembourg, p. 15

Process / Procedure

A series of actions or operations that make gradual changes leading towards a particular result

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Program

A group of related surveys

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Property

A peculiarity common to all members of an object class

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Public disclosure

Public disclosure refers to the act of making information or data readily accessible and available to all interested individuals and institutions. Some examples of the different forms that public disclosure may

take include: verbal or written statements released to a public forum, to the news media, or to the general public; publication in an official bulletin, gazette, report, or stand-alone document; and information posted on a website.

Source: "Code of Good Practices on Transparency in Monetary and Financial Policies", Part 1- Introduction; approved by the IMF Executive Board on July 24, 2000
Hyperlink: http://www.imf.org/external/np/mae/mft/sup/part1.htm#appendix_III

Public Use Files (PUF)

Public Use Files (PUFs) are available to anyone agreeing to respect a core set of easy-to-meet conditions. Such conditions relate to *what* can be done with the data (e.g. the data cannot be sold), not to access to the data. In some cases PUFs are disseminated with no conditions; often being made available on-line. These data are made easily accessible because the risk of identifying individual respondents is considered minimal. Minimising the risk of disclosure involves eliminating all content that can identify respondents directly—for instance, names, addresses and telephone numbers. In addition this requires purging relevant indirect identifiers from the microdata file.

Source: Olivier Dupriez and Ernie Boyko. 2010, "Dissemination of Microdata Files; Principles, Procedures and Practices" IHSN Working Paper No. 005, August 2010, p. 6
http://www.ihsn.org/home/index.php?q=focus/dissemination-_microdata-files-principles-_procedures-and-practices

Raw microdata files

Raw microdata files contain all replies by each respondent obtained immediately after data entry.

Source: Olivier Dupriez and Ernie Boyko. 2010, "Dissemination of Microdata Files; Principles, Procedures and Practices" IHSN Working Paper No. 005, August 2010, p. 5
http://www.ihsn.org/home/index.php?q=focus/dissemination-_microdata-files-principles-_procedures-and-practices

Remote Access Data Laboratory (RADL) – Australia

The Remote Access Data Laboratory (RADL) is a web-based tool that allows authorised users to access detailed microdata that is stored within the Australian Bureau of Statistics (ABS) secure environment. Built-in automatic checks prevent large-scale release of unit record information, thus maintaining confidentiality of data providers as outlined in Australian legislation.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 50
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Remote access

Remote Access Facilities

These are facilities that provide researchers with the ability to produce statistical outputs from microdata through computer networks without researchers actually 'seeing' the microdata. The microdata itself does not leave the National Statistical Office. Remote Access Facilities may be of two types.

(a) Remote execution where a researcher submits a programme and receives the output later by email.

(b) Remote facilities where the researcher performs the analysis and can immediately see the answer on the screen.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 107
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Remote Data Access

This consists of providing users with access to web-based data tabulation and analysis software, with no possibility of users downloading datasets or generating tables that would reveal individual or small numbers of records.

Source: Olivier Dupriez and Ernie Boyko. 2010, "Dissemination of Microdata Files; Principles, Procedures and Practices" IHSN Working Paper No. 005, August 2010, p. 8
http://www.ihsn.org/home/index.php?q=focus/dissemination-_microdata-files-principles-_procedures-and-practices

Remote data access (RDA) - Canada

Remote data access (RDA) is a mode of indirect access to confidential microdata through which researchers submit their own computer programs via the Internet to Statistics Canada, where they are run by Statistics Canada staff on the internal unscreened microdata. The results are then vetted for confidentiality and sent back to the researcher.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 47
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Research community

Although this mainly refers to people working in research institutions such as universities, it also includes researchers working in government agencies, NGOs, international agencies and the private sector. Some countries may want to define the research community more narrowly and only include those working in research institutions.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 106
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Research Data Centres (RDC) – Canada

Starting in 2000, Statistics Canada, in partnership with participating Canadian universities, the Social Sciences and Humanities Research Council and the Canadian Foundation for Innovation, established a network of Research Data Centres (RDC) in Canadian universities. These centres are enclaves of Statistics Canada, within which researchers have access to household survey data in an environment that respects Statistics Canada's requirements for security and confidentiality. There are currently 15 RDC locations

across the country, plus a federal RDC in Ottawa used by statistical researchers in federal government departments.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 56
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Research Data Centres (RDC) - United States

Research Data Centres (RDCs) offer qualified researchers restricted access to confidential economic and demographic data collected by the U.S. Census Bureau in its surveys and censuses. All projects must offer benefits to U.S. Census Bureau programmes. These projects are carried out at U.S. Census Bureau headquarters, or at one of eight other secure locations around the U.S.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 59
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Research files

See Licensed files

Risk avoidance

This approach tries to eliminate all risks. In the case of microdata confidentiality, it requires the confidentiality of the data to be absolute, not only in its own right, but in association with other available data.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 107
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Risk management

Within the constraints provided by legislation, it involves identification of the risks and managing them in accordance with their significance (impact) and their likelihood. More effort is put into managing the high impact, strong likelihood risks. Microdata confidentiality may not be absolute when considered in association with other data. Confidentiality could be considered in association with other means of reducing the risk.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 107
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Security

Specific attention should be devoted to supporting the use of techniques and instruments to guarantee the integrity and security of research data. With regard to guaranteeing the integrity of a data set, every effort

should be made to ensure the completeness of data and absence of errors. With regard to security, the data, along with relevant meta-data and descriptions, should be protected against intentional or unintentional loss, destruction, modification and unauthorised access in conformity with explicit security protocols. Data sets and the equipment on which they are stored should be protected as well from environmental hazards such as heat, dust, electrical surges, magnetism, and electrostatic discharges.

Source: OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding,* Paris, p. 20

Special Data Dissemination Standard (SDDS)

The Special Data Dissemination Standard (SDDS) was established by the International Monetary Fund (IMF) to guide members that have, or that might seek, access to international capital markets in the provision of their economic and financial data to the public. Subscription to the SDDS was opened in early April 1996.

Context: The SDDS identifies four dimensions of data dissemination:

a) The data: coverage, periodicity, and timeliness;

b) Access by the public;

c) Integrity of the disseminated data; and

d) Quality of the disseminated data.

The SDDS prescribes that subscribing members provide a summary description of methodology for each data category, including statements of major differences from international guidelines. The term "methodology" is used in the SDDS in a broad sense to cover the aspects of analytical framework, concepts, definitions, classifications, accounting conventions, sources of data, and compilation practices.

Source: International Monetary Fund (IMF), "Guide to the Data Dissemination Standards, Module 1: The Special Data Dissemination Standard", Washington, May 1996
Hyperlink: http://dsbb.imf.org/sddsindex.htm

Statistical catalogue

A code list of statistical indicators, metadata, questionnaires, tables and other defined elements of a statistical information system

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Statistical data - 1

Data that are collected and/or generated by statistics in process of statistical observations or statistical data processing.

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000

Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Statistical data - 2

Statistical data refers to data from a survey or administrative source used to produce statistics

Source: Measuring the Non-Observed Economy: A Handbook, OECD, IMF, ILO, Interstate Statistical Committee of the Commonwealth of Independent States, 2002, Annex 2, Glossary.

Hyperlink: http://www.oecd.org/dataoecd/9/20/1963116.pdf

Statistical Data and Metadata Exchange (SDMX)

Statistical Data and Metadata Exchange (SDMX) is an initiative sponsored by BIS, ECB, Eurostat, IMF, OECD, UN and World Bank to address standardization of the exchange of statistical information.
Source: Statistical Data and Metadata Exchange (SDMX) – BIS, ECB, Eurostat, IBRD, IMF and OECD – Metadata Common Vocabulary, Release 1, December 2003.
Hyperlink: www.sdmx.org

Statistical disclosure

Statistical disclosure is said to take place, if the dissemination of a statistics enables the external user of the data to obtain a better estimate for a confidential piece of information than would be possible without it.

Source: Eurostat, "Manual on disclosure control methods", Office for Official Publications of the European Communities, Luxembourg, 1996, p. 7

Statistical macrodata

An observation data gained by a purposeful aggregation of statistical microdata conforming to statistical methodology.

Context: Macrodata is data derived from microdata by statistics on groups or aggregates, such as counts, means, or frequencies. (Survey Design and Statistical Methodology Metadata, Software and Standards Management Branch, Systems Support Division, United States Bureau of the Census, Washington D.C., August 1998, Section 3.4.4, page 39)

Source: Economic Commission for Europe of the United Nations (UNECE), "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000
Hyperlink: http://www.unece.org/stats/publications/53metadaterminology.pdf

Statistical macrodata

An observation data gained by a purposeful aggregation of statistical microdata conforming to statistical methodology.

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Statistical macrodatabase

The database containing statistical macrodata

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Statistical metadata -1

Metadata describing statistical data

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Statistical metadata -2

Metadata provide information on data and about processes of producing and using data. Metadata describe statistical data and - to some extent - processes and tools involved in the production and usage of statistical data (UNECE, "Guidelines for the Modelling of Statistical Data and Metadata", 1995).

Source: United Nations Statistical Commission and Economic Commission for Europe of the United Nations (UNECE), "Guidelines for the Modelling of Statistical Data and Metadata", Conference of European Statisticians, Methodological material, United Nations, Geneva, 1995
Hyperlink: http://www.unece.org/stats/publications/metadatamodeling.pdf

Statistical microdata -1

An observation data collected on an individual object - statistical unit.

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Statistical microdata -2

An observation data collected on an individual object statistical unit.

Context: Microdata is data on the characteristics of units of a population, such as individuals, households, or establishments, collected by a census, survey, or experiment. (Survey Design and Statistical Methodology Metadata, Software and Standards Management Branch, Systems Support Division, United States Bureau of the Census, Washington D.C., August 1998, Section 3.4.4, page 39)

Source: Economic Commission for Europe of the United Nations (UNECE), "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000
Hyperlink: http://www.unece.org/stats/publications/53metadaterminology.pdf

Statistical microdatabase

The database containing statistical microdata

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Statistical purposes

It is particularly important to make a distinction between statistical and administrative uses. In the case of statistical use, individual data are used as an input to derive statistics that refer to a group of persons or legal entities. It may also incorporate support for other activities within a NSO (e.g. sample selection off a business register). Administrative uses concern decisions about a particular person or legal entity which may bring benefit or harm to the individual. The statistics referred to above include statistical aggregates, statistical distributions, parameters for models and other forms of statistical analysis that may refer to groups of individuals or organizations without identifying them. Microdata used for research is consistent with statistical purposes if it is being used to produce the type of statistics referred to in the previous paragraph.

Source: UNECE, "Managing Statistical Confidentiality & Microdata Access - Principles and Guidelines of Good Practice", 2007, p. 106
*Hyperlink:*http://www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Statistical standard

A statistical standard provides a comprehensive set of guidelines for surveys and administrative sources collecting information on a particular topic.

Components of a standard include:

- definition(s)
- statistical units
- classification(s)
- coding process(es)
- questionnaire module(s)
- output categories

Context: The use of statistical standards permits the repeated collection of statistics on a consistent basis. They also enable the integration of data over time and across different data sources, allowing the use of data beyond the immediate purpose for which it was produced. Standards also reduce the resource requirements associated with many aspects of survey development and maintenance.

Source: Statistics New Zealand," Classifications and Standards"; unpublished on paper
Hyperlink: http://www.stats.govt.nz/domino/external/web/prod_serv.nsf/092edeb76ed5aa6bcc256afe0081d84e/35b11e7066c13db1cc256ca5006f44e4?OpenDocument

Supplementary data

In SDMX, "Supplementary Data" refers to a description of data not routinely disseminated that are made available to users upon request. It may include customized tabulations that can be provided (perhaps for a fee) to meet specific requests. Also include information on procedures for obtaining these supplementary data.

Source: Statistics Canada, "Statistics Canada Quality Guidelines", 3rd edition, October 1998, page 59
Hyperlink: http://www.statcan.ca/english/freepub/12-539-XIE/12-539-XIE.pdf

Sustainability

Due consideration should be given to the sustainability of access to publicly funded research data as a key element of the research infrastructure. This means taking administrative responsibility for the measures to guarantee permanent access to data that have been determined to require long-term retention. This can be a difficult task, given that most research projects, and the public funding provided, have a limited duration, whereas ensuring access to the data produced is a long-term undertaking. Research funding agencies and research institutions, therefore, should consider the long-term preservation of data at the outset of each new project, and in particular, determine the most appropriate archival facilities for the data.

Source: OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, Paris, p. 22

Statistical unit for macrodata

A statistical unit which is a carrier or a supplier of statistical macrodata in statistical system

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Statistical unit for microdata

A statistical unit which is a carrier or a supplier of statistical microdata in statistical system

Source: UNECE, Conference of European Statisticians Statistical Standards and Studies – No. 53, "Terminology on Statistical Metadata", Geneva, 2000
Hyper link: http://www.unece.org/stats/publications/53metadaterminology.pdf

Timeliness

Speed of dissemination of the data - i.e., the lapse of time between the end of a reference period (or a reference date) and dissemination of the data.

Context: In SDMX, "Timeliness and Punctuality" is a single entity. Timeliness refers to the speed of dissemination of the data - i.e., the lapse of time between the end of a reference period (or a reference date) and dissemination of the data. It reflects many factors, including some that are related to institutional arrangements, such as the preparation of accompanying commentary and printing.

Punctuality refers to the possible time lag existing between the actual delivery date of data and the target date when it should have been delivered, for instance, with reference to dates announced in some official release calendar or previously agreed among partners.

Source: International Monetary Fund (IMF), "Guide to the Data Dissemination Standards, Module 1: The Special Data Dissemination Standard", Washington, May 1996
Hyperlink: http://www.imf.org/external/bopage/pdf/mar2000.pdf

Transparency -1

Transparency refers to an environment in which the objectives of policy, its legal, institutional, and economic framework, policy decisions and their rationale, data and information related to monetary and financial policies, and the terms of agencies' accountability, are provided to the public in a comprehensible, accessible, and timely manner.

Source: Code of Good Practices on Transparency in Monetary and Financial Policies, Part 1—Introduction, Approved by the IMF Executive Board on July 24, 2000.

Hyperlink: http://www.imf.org/external/np/mae/mft/sup/part1.htm#appendix_III

Transparency -2

Information on research data and data-producing organisations, documentation on the data and specifications of conditions attached to the use of these data should be internationally available in a transparent way, ideally through the Internet. Lack of visibility of existing research data resources and future data collection poses serious obstacles to access. Factors to consider in ensuring transparency include:

• Information on data-producing organisations and their holdings, documentation on available data sets and conditions of use should be easy to find on the Internet.

• Research organisations and government research agencies should actively disseminate information on research data policies to individual researchers, academic associations, universities and other stakeholders in the publicly funded research process.

• Whenever relevant, all members of the various research communities should assist in establishing agreements on standards for cataloguing data. The application of existing standards should be considered, whenever appropriate, in order to avoid placing additional burdens on research resources and work loads of researchers and their institutions.

• Information on data management and access conditions should be communicated among data archives and data producing institutions, so that best practices can be shared.

Source: OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding,* Paris, p. 15.