

Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS)

11-13 March, 2009

Do conceptual systems improve concepts effectiveness?

Prepared by Luísa Saraiva
Statistics Portugal

I. Digging below the surface...

1. One of the multiple challenges that face Statistical Offices and International Organizations dealing with statistics is the need to invest on improving management and interchange of data and metadata.
2. Measuring performance, quality, goals and the fulfilment of international standards requires each time more accuracy in the collecting, treatment, sharing and dissemination of statistical information.
3. This state of the art leads us to believe that the efforts in progress towards the adoption of a terminological methodology in the harmonization of concepts and the construction of conceptual systems in Statistics Portugal as a means of organizing and representing knowledge provide efficiency to data and metadata.

II ...concerning concepts subsystem in Statistics Portugal, ...

4. The implementation of an integrated concepts management system to ensure the adequate technical coordination, promote the dissemination, the updating and user access to metadata have been Statistics Portugal organizational priorities since 1990, clearly conveyed by the general guidelines laid down by the Statistical Council on National Statistical Activity.

5. The concepts subsystem of Statistics Portugal is the repository of the concepts used in surveys and administrative sources related to the national Statistical System, which are periodically sent to the Statistical Council for further analysis, evaluation and approval, becoming afterwards of compulsory application to official statistical data.

6. The concepts subsystem architecture consists on a management database and two WEB applications directly available for consultation on the home page of the official Statistics Portugal website as well as on its intranet, classified by subject area and organised in glossaries in accordance to the themes adopted for the dissemination of data.

7. Core attributes of the entity “concept” are code, term/designation, definition, notes on the definition and source, which function is to convey its essential features. There are however other attributes also relevant for the management of the system: status and validity (proposed and in use, being this one approved or not by the Statistical Council) particularly important to document the entire life cycle of surveys and data, specifically the breaking of data series; thematic area and glossary, issuing the structuring of subject areas; formula that allows another representation of the concept; synonymy and homonymy establishing semantic relationships with other concepts; history recording concept’s lifetime and last but not the least methodological documents, classifications and variables showing and allowing linkages to the correspondent subsystems and concepts usages.

8. A list of abbreviations and acronyms employed in surveys documentation is also available at the home page of the concepts subsystem as well as the access to other concept systems like those of OECD and Eurostat.

9. The information retrieval is accomplished by means of a browser that enables simple search by alphabetical order in each glossary and advanced search by the combination of more than one search criterion such as code, glossary, source, word, term and status among others.

10. Future improvements are under analysis related to the recording of concepts, the implementation of new attributes and conceptual systems in order to restructure the database and web applications concerning concepts information retrieval to users in general.

III. ...grasping the meaning, to occupy the right place, ...

11. A conceptual system is a “structured set of concepts established according to the relations between them, each concept being determined by its position in the set”, which applied to statistics thematic domains, aim to improve the accuracy of data and metadata as well as strengthen their consistency.

12. The identification of concepts characteristics show the similarities and delimitations between them, being essential to the evaluation of the connections supporting their relations and the revealing of gaps and inadequacies: concepts do not exist as isolated units of knowledge, but always in relation to other(s)!

13. Concepts relations are hierarchic and associative. Hierarchic relations organize concepts into levels where the superordinate concept is subdivided into at least one subordinate concept: subordinate concepts at the same level and having the same criterion of subdivision are called coordinate concepts and share a dimension. Concepts are not superordinate, subordinate or coordinate, on their own, but in relation to each other in a hierarchy. Hierarchic relations may be either monodimensional or multidimensional.

14. There are two types of hierarchical relations: generic and partitive. Generic relations exist when the subordinate concept includes the characteristics of the superordinate plus at least one additional delimiting characteristic. Partitive relations exist when the superordinate concept represents a whole, while the subordinate concepts represent parts of that whole that come together to form the whole: the superordinate concept is called the comprehensive concept and the subordinate concept is called the partitive concept. Subordinate concepts at the same level and sharing the same dimension are also called coordinate concepts.

15. Associative relations establish a thematic connection between concepts based on their proximity in space, time, cause or effect beyond others. There are many kinds of associative relations.

IV. ... tying designations and definitions to their respective concepts!

16. There are three essential conditions to build a conceptual system, taking into account that it is unfinished by nature:

- the exhaustive understanding of conceptualization that structures knowledge in a subject area
- the acquaintance with the objectives and expectations of real and potential users
- the capacity of becoming easily adjusted to the changes, evolution and expansion that will inevitably occur.

17. In what concerns methodology, several stages are supposed to be accomplished all along the process, some of them simultaneously:

- the gathering of unstructured but thematically related concepts within a specialized domain,
- the consulting of every related and available sources of information (legislation, guidelines, terminological bases, dictionaries...)
- the allocation of the concepts within the concept system, grouping them into meaningful units
- the attribution of the designations and the formulation of the definitions for each one, bearing in mind concepts relations.

18. The concepts selection, the terms attribution and the definitions formulation represent the way the model of the concepts structure is achieved. The quality of most terminological products will be determined by the quality of the definitions.

IV. How far leads the path?

19. Concepts systems play a pivotal role by modelling concept structures based on specialized knowledge of a field, clarifying relations between concepts, settling up the basis for a uniform and standardized terminology regarding data and metadata interchange and accountability, facilitating comparative analysis of concepts and designations as well as the writing of definitions.

20. The increasing share of knowledge claims for good practices on ensuring consistency in the adoption of statistical standards and concepts across the countries, aiming to reduce ambiguity, avoid misunderstandings and clarify their interconnection to respondents, decision-makers and public in general.

21. The convergence of internationally shared conceptual statistical systems is an undertaking of vital importance in a growing global world.

21. Let's follow the path and take the risk, shall we?

References

OECD (2007), *Glossary of statistical terms*, Paris (Pdf version) on <http://stats.oecd.org/glossary>.

INE (2007), *Linhas gerais da actividade estatística nacional e respectivas prioridades, 2008 – 2012*. Projecto DOct/2102/CSE/PCD

ISO (2000), *Terminology work - principles and methods*, ISO/FDIS 704

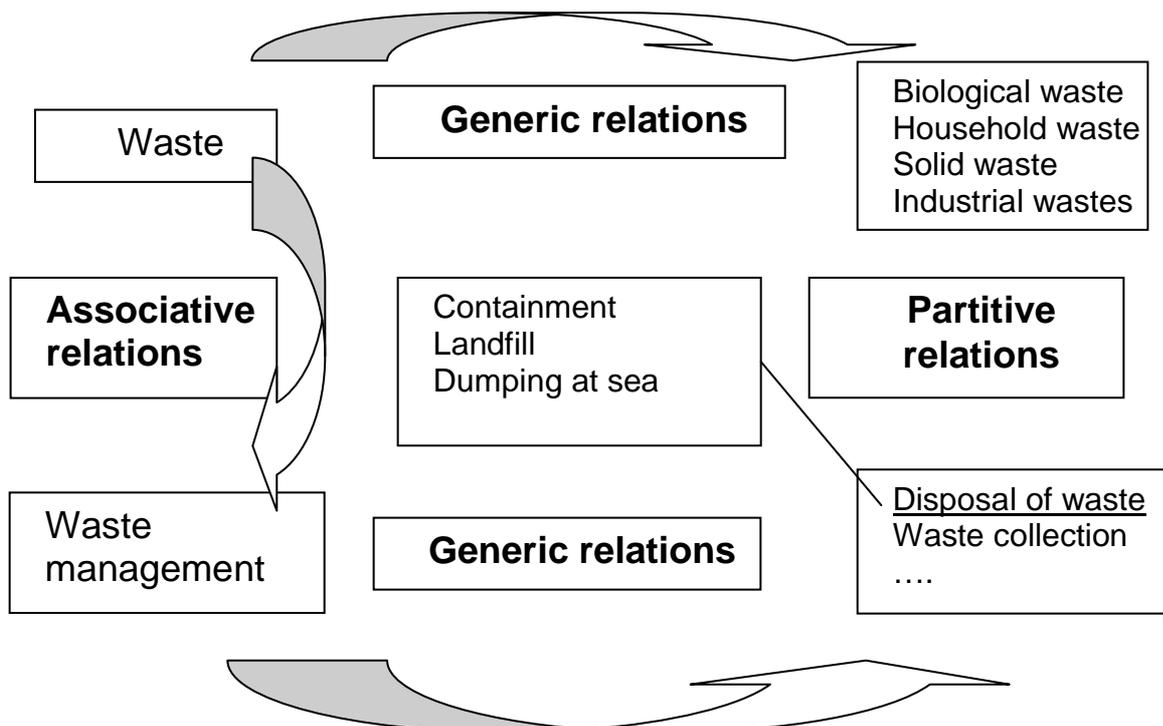
Morgado, I. and Isfan, M. (2008), "Case study – Statistics Portugal", paper presented at METIS, Luxembourg.

Sager, J (2000) *A practical course in terminology processing*. Amsterdam, John Benjamins

Saraiva, L (2007), The role of conceptual systems in the organizing of knowledge (an exercise within the Environment domain), paper presented at the 56th Session of the ISI, Portugal

Annex 1- Example of part of a conceptual system
Statistical domain *Environment*

| Unstructured concepts | Structured concepts |
|-----------------------|-------------------------|
| Environment | 1 Environment |
| ... | |
| Biological waste | 4 Waste |
| Containment | 4.1 Biological waste |
| Disposal of waste | 4.2 Household waste |
| Dumping at sea | 4.3 Solid waste |
| Household waste | 4.4 Industrial wastes |
| Industrial wastes | 4.5 Waste management |
| Landfill | 4.5.1 Disposal of waste |
| Solid waste | 4.5.1.1 Containment |
| Waste | 4.5.1.2 Landfill |
| Waste collection | 4.5.1.3 Dumping at sea |
| Waste management | 4.5.2 Waste collection |
| ... | ... |



Annex 2 - Concepts extracted from the OECD Glossary of statistical terms

Waste Waste refers to materials that are not prime products (that is, products produced for the market) for which the generator has no further use in terms of his/her own purposes of production, transformation or consumption, and of which he/she wants to dispose. *Context:* Wastes may be generated during the extraction of raw materials, the processing of raw materials into intermediate and final products, the consumption of final products, and other human activities. Residuals recycled or reused at the place of generation are excluded.

Biological waste Biological waste is waste containing mostly natural organic materials (remains of plants, animal excrement, biological sludge from waste-water treatment plants and so forth)

Household waste Household waste refers to waste material usually generated in the residential environment. Waste with similar characteristics may be generated in other economic activities and can thus be treated and disposed of together with household waste

Industrial wastes Industrial wastes are liquid, solid and gaseous wastes originating from the manufacture of specific products

Solid waste Solid waste is useless and sometimes hazardous material with low liquid content. Solid wastes include municipal garbage, industrial and commercial waste, sewage sludge, wastes resulting from agricultural and animal husbandry operations and other connected activities, demolition wastes and mining residues

Waste collection Waste collection is the collection and transport of waste to the place of treatment or discharge by municipal services or similar institutions, or by public or private corporations, specialized enterprises or general government. Collection of municipal waste may be selective, that is to say, carried out for a specific type of product, or undifferentiated, in other words, covering all kinds of waste at the same time

Waste management The characteristic activities of waste management include: (a) collection, transport, treatment and disposal of waste, (b) control, monitoring and regulation of the production, collection, transport, treatment and disposal of waste, and (c) prevention of waste production through in—process modifications, reuse and recycling

Disposal of waste Disposal of waste refers to waste elimination techniques comprising landfills, containment, underground disposal, dumping at sea and all other disposal methods

Containment Containment is the retention of hazardous material so as to ensure that it is effectively prevented from dispersing into the environment, or released only at an acceptable level. Containment may occur in specially built containment spaces
Source: Glossary of Environment Statistics, Studies in Methods, Series F, No. 67, United Nations, New York, 1997

Landfill Landfill refers to the final placement of waste in or on the land in a controlled or uncontrolled way according to different sanitary, environmental protection and other safety requirements.

Dumping at sea Dumping at sea is the disposal of hazardous and non-hazardous substances in the open sea

Note: The source mentioned in all the concepts records is the Glossary of Environment Statistics, Studies in Methods, Series F, No. 67, United Nations, New York, 1997