UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE (ECE)
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Bratislava, Slovakia, 18-20 April 2005)

Topic (iv): Review and follow-up to the activities of the Conference of European Statisticians

SYSTEM OF DISSEMINATION (STATWEB 4.0)
OF STATISTICAL DATA AND METADATA AT STATISTICS NETHERLANDS

Supporting Paper

Submitted by Statistics Netherlands[1]

Summary

I. INTRODUCTION

1.   This paper describes the process Statistics Netherlands (SN) is undertaking now to realize a better and more consistent way of publishing statistical data and metadata. Achieving coordination of all metadata used in the statistical process is an important aim, that also will benefit the ambition of being a Bureau of Standards.

II. COORDINATION

2.   Implementation of coordination can be reached by offering tools, instead of issuing rules. The tools offered, however, should influence all phases of statistical production. Recently both vertical and horizontal integration of the statistical production focused on the need to coordinate all phases of statistical production; input, throughput and output. The first step in the process was based on the experiences of the ABS, resulting in the use of a classification-server in SN, aiming at coordination of output.

3.   The results were promising, but because of acceptance many variants on standards had to be accepted in the Classification-server, reducing its effectiveness. Based on the Cristal model, a formalized translation and extension of core meta-statistical concepts that make up the structure of databases and tables, it will be possible to facilitate the transformation of poorly and intuitively structured data into better considered and more disciplined structures

---

[1] Prepared by Albert Jacques (ajcs@cbs.nl).

III.   METADATA

4.   Based on the flexible underlying Cristalmodel and the results of the Neuchâtel Group a new kind of metadata has been developed. The metadata of the statistical output are grouped into four kinds of metadata; conceptual, process, quality and technical metadata.

5.   The information about statistical data (metadata) can be determined if it does not contain the statistical result. These metadata are called conceptual metadata in the Dutch metadata model. Different processes can lead to different data, with the same conceptual metadata. To distinguish these different results, process metadata and quality metadata are needed.

IV.   THE NEW DISSEMINATION DATABASE STATLINE (STATWEB 4.0)

6.   In the present dissemination database StatLine (StatWeb 3) tables are built as independent units. Identical variables are placed in more than one table, risking that for identical indicators different data can be published and also different metadata can exist for identical indicators in different tables. In the new data model StatWeb 4.0 both statistical information and meta-information can be reused easily. It is also possible to make statistical metadata available to third parties without statistical information, which fits exactly in the ambition of Statistics Netherlands being a Bureau of Standards.

V.   LEVEL OF COORDINATION IN METADATA

7.   All metadata explicitly indicated as such by the Board of Directors of SN are coordinated metadata. Coordinated metadata can be a National Standard or are seen as other coordinated metadata. Coordinated metadata are always shared. All other metadata are uncoordinated metadata, but will be included in the metadata server, be it with an expiry date. Uncoordinated metadata can be shared (used in more than one process) or be non-shared. The use of non-shared uncoordinated metadata is reserved to single users only.

VI.   RULES OF IMPLEMENTATION FOR COORDINATED METADATA

8.   Coordinated metadata define protected concepts. These concepts may not be used by SN within the same context but with different meaning. When confidentiality and quality allow it, SN makes publications according to the national standards. The national standards always prevail over the coordinated standards, they in turn prevail over the uncoordinated standards. When aggregating certain categories in a classification, a new variant of the classification is stored along with the original one. As such it is not necessary to give these variants names, but it is possible when necessary. Uncoordinated metadata will be included in the metaserver, because when SN decides so, specific requirements of users of statistical information have to be recognized. Often those are additional publications apart from the regular ones, which are based on coordinated metadata. These uncoordinated metadata will be placed on a 'tolerated list', having an expiry date, which is at the maximum two years later. The temporary metadata still not coordinated are placed on the 'tolerated list'. This list explains which metadata are allowed to be used for new tables or updates of existing tables only. Old metadata, used for tables no longer subject for updates may be used longer, having an expiry date in the past. Uncoordinated metadata can either be shared or non-shared.

VII.   IMPLEMENTATION IN SN

9.   Implementation, in fact, is a combination of two processes. The first one, the implementation of StatWeb 4.0; the second one the integration of the present C-server and the other metaservers in StatWeb 4.0. The most important differences between StatWeb 3 and Statweb 4.0 are: (a) a split-out of tables into cubelets, and (b) also splitting out the metadata lacing them together with the tables, making links to a central place where all metadata are placed. To realize this, the first step is a conversion; all tables from StatWeb 3 have to be converted into cubelets into StatWeb 4.0. Statistical Coordination made up standards for, activity classification, sex, age, size-class by turnover and region. This standard will be used in the conversion. For other classifications and concepts the present central metadata will be used. For the first six months after

conversion, tables will be imputed both in StatWeb 3 and StatWeb 4.0. A new administration for the transactions will be set up. A communication plan will be written to inform users about the process.

VIII.   CONCLUSIONS

10.   By integrating metadata in the dissemination system Stateline, the obligation to make use of this metadata is obvious, publication is impossible otherwise. Not only classifications are metadata, but also definitions and concepts. The Cristal model offers a useful instrument for this integration. StatWeb 4.0 is seen as an important step towards national and international coordination of statistical information produced by Statistics Netherlands. Combined with the ambition of being a Bureau of Standards, Statistics Netherlands hopes to play an important role in the coordination of the complete statistical process and radiate the importance and benefits of coordination.