

Distr.  
GENERAL

CES/AC.71/2003/7 (Summary)  
3 December 2002

Original: ENGLISH

**UNITED NATIONS STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE  
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**  
(Geneva, 17-19 February 2003)

Topic II: Impact of technical measures and standards on data quality

**NEW DEVELOPMENTS IN INFORMATION PROCESSING  
AT STATISTICS NETHERLANDS**

**Invited paper**

Submitted by Statistics Netherlands<sup>1</sup>

**Summary**

1. Two years ago, Statistics Netherlands drastically reshaped its organizational structure and formulated its strategy for the coming years. Three elements in this strategy were of special interest to its information technology department. First of all, the organizational structure of Statistics Netherlands was changed from a statistics-oriented structure (also known as stovepipes) to a process-oriented structure. Second, focus was shifting from carrying out surveys for data collection to linking to administrative registers which contain large collections of data. Third, attention was shifting from a quantity-oriented output process to a more quality-aware output process. In this paper we describe some of the new developments that were introduced by the information technology department to support the new strategy. In addition, we indicate how these developments influence the quality of data in the production processes and what kind of impact they have on the quality of data being published.

2. One of the new developments we like to mention here is the increasing use of *data warehouse technology* in statistical processes. We found that data warehouses are suited for storing large collections of data from registers. In this way the data can be made easily accessible for statisticians using standardized tools for on-line analytical processing. Although this may sound simple, in reality it is not. Designing, loading and maintaining such a data warehouse is a complex task. But once the data is loaded and cubes are defined, statisticians can explore the data better and more flexible than they did before. In fact, they happened to see things they never saw before. How this affects data quality in general is not yet known, but as statisticians acquire more powerful capabilities to analyse their data, we expect them to be able to be more complete in their results, which is one of the dimensions of data quality.

---

<sup>1</sup> Prepared by Olav ten Bosch (obos@cbs.nl).

3. Another development we like to mention is the introduction of *metadata repositories*. These metadata repositories contain standardized information like statistical classifications, variables and mappings. Although the introduction of metadata repositories has a long history, the new process-oriented structure of Statistics Netherlands made these metadata repositories more feasible to actually be used in new systems. In addition, they were opened up via component interfaces, so that they could be used easily from any information system in the organization. We found that a very strict metadata policy (enforced use of metadata repositories) was not useful. On the other hand, we found that a more relaxed metadata policy (please use it when possible) enhances the comparability between different statistical (sub)processes and therefore has a positive impact on the overall coherence of statistics.

4. Another development, which was introduced quite recently, is the introduction of the *RUP* (Rational Unified Process) in the information technology department. Although the introduction is still in a preliminary phase, two pilot projects were already carried out using this method for developing software. Especially the use of a pre-defined framework for specifying requirements for information systems proved to be valuable for both end-users as well as system analysts. Furthermore, the information systems that were designed and built in the pilot projects contained very few bugs, which is an indication of their quality. In general, we expect the introduction of this method to have a positive impact on the quality of information system development and maintenance, and therefore on the quality of the statistics process as a whole.

5. The last issue we describe here is related to the statistical database *StatLine*. In about 10 years, this system evolved from a quite simple, table-generating program into a flexible, open and highly customizable multidimensional database that enables end-users to interactively browse its contents via the internet. Currently, StatLine contains *all* of the statistical results published by Statistics Netherlands. A very useful feature of StatLine is the ability to point to any subset of its data via one simple concise hyperlink. This facility has been used for about two years now and it proved to be surprisingly useful. In the near future this mechanism will be extended to use standard web service technology, transforming it into a *statistical web service*. This service will offer external systems automated access to statistical content via the internet. Customers may use this service to automatically check certain statistical indicators at regular intervals. Another improvement is the incorporation of specialized GIS technology in StatLine. With this technology, users can actually browse the database via a geographical entry. Both the statistical web service as well as the geographical entry improve the accessibility and clarity of the data and therefore contribute to the quality of the statistical data being published.

-----