Distr.
GENERAL

CES/AC.71/2003/8 (Summary)
6 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

ENHANCING DATA QUALITY THROUGH DATABASE INTEGRATION AT OECD

Invited paper

Submitted by OECD[1]

Summary

1. The most important strategic programme of the OECD Statistics Directorate is the reinforcement of the quality and efficiency of the Organisation's statistical work. It is based on two projects: the establishment of a quality framework and the implementation of a new statistical infrastructure. This paper describes how the new infrastructure supports the implementation of the quality framework.

2. The OECD quality framework has three broad aims. Firstly, it provides a systematic mechanism for ongoing identification and resolution of quality problems. Secondly, it gives greatly increased transparency to the processes used by the OECD to assure quality. Thirdly, it reinforces the political role of the OECD in the context of an information society. The framework comprises: a definition of quality and its dimensions; a procedure for assuring the quality of proposed new statistical activities; a procedure for evaluating the quality of existing statistical activities on a regular basis; and quality guidelines. The Organisation developed the quality framework in 2002 and plans to implement it in the course of 2003. The work on quality has been conducted in parallel with the development of new statistical infrastructures.

3. The OECD has defined quality in terms of the following eight dimensions: relevance, accuracy, credibility, timeliness, punctuality, accessibility, interpretability and coherence. The quality characteristics of most importance depend on user perspectives, needs and priorities, which vary across groups of users. In addition cost-efficiency is a factor that must be taken into account. The trade-offs between the various ways in which quality can be improved and their costs must be considered.

[1] Prepared by Gérard Salou (gerard.salou@oecd.org).

4. The new statistical infrastructures provide new statistical tools and support decentralised statistical activities in order to increase the incentives for those activities to adopt the quality guidelines and to share common systems and practices. The tools are tightly integrated together and cover the following elements: centralized detailed information on each of the OECD statistical activity, a common electronic glossary of statistical terms for the harmonisation of terminology and concepts, a set of common metadata items, a new dataset comprising the data series that are the most commonly used across the Organisation, a complete data catalogue and a common data browser to navigate the various OECD datasets. Those tools contribute to the reinforcement of all quality dimensions and of efficiency of statistical and analytical tasks. The following paragraphs describe the particular impact on accessibility, interpretability and coherence.

5. The most obvious quality dimension immediately improved by the new statistical infrastructure is accessibility. It is improved by the central information on individual statistical activities and by the use of state-of-the-art technology. The central information is updated by all statistical areas of the decentralized OECD statistical system. It is then used by analysts in the Organisation as a major discovery tool for navigating the 95 OECD datasets. Another discovery tool is the Glossary of statistical terms that can be searched in a sophisticated way. The Glossary is linked to data elements of the data warehouse, in particular for the most frequently used series which we call Reference Series. In addition, the general interface on OECD.Stat permits to search on all common metadata items. Accessibility for outside users will be much improved when the Web Services components that are being developed are implemented on a production basis

6. Interpretability is improved by central tools for data documentation and by improved accessibility to metadata. The Glossary, central metadata items and a common set of metadata items are the main instruments. Coherence of metadata is also an important factor for improving interpretability.

7. Coherence has four important sub-dimensions: within a dataset, across datasets, across countries and over time. Coherence within datasets is improved in OECD.Stat by the fact that data are related to the central glossary which forces data concepts to be in-line with their official definition and other attributes, whether internationally agreed or not. Coherence across datasets is improved by the data confrontation permitted by the data warehouse in the improved accessibility and interpretability of data.

8. In technical terms, OECD.Stat is developed around Microsoft tools. MS SQL Server 2000 and its OLAP component are used as central data and metadata repository. Data flows between individual production areas and the central systems use an internal XML format. An Excel Wizard has been developed to permit easy access to Reference Series and a web-based interface is on development as general user interface to the entire system. Finally, web services are being developed for delivering data outside the OECD.

- - - - -