

Distr.
GENERAL

CES/AC.71/2003/16 (Summary)
4 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

DATA LIFE-CYCLE OBJECT MODEL FOR STATISTICAL INFORMATION SYSTEMS

Supporting paper

Submitted by Eurostat¹

Summary

1. Statistical information systems deal with the processing of data sets collected from data providers to produce statistics. These data sets are processed in the form of data flows, the stream of which is governed by work flows. At each step of the work flow, the data sets are in a given state and can be aggregated in larger or more abstract data sets. The paper presents an object-oriented meta-model, which depicts several aspects of the data life cycle of data sets in statistical information systems. Models are presented in the UML notation.
2. As any statistical institute, Eurostat collects raw data (either micro-data or aggregated data) from many sources, which are then validated and processed to produce value-added statistical information to be used by decision-makers, politicians, journalists, researchers or citizens. Data are received and processed as data sets, which go through various steps and states and get filtered and aggregated before they can be published as useful new information. This is called the Data Life Cycle or CVD process (for Cycle de Vie des Données) at Eurostat.
3. The paper presents an object-oriented meta-model, which depicts various aspects of the data life cycle of data sets in statistical production systems. The models are derived from Eurostat's experience but are by no means specific to this EU agency and can serve as meta-models or reference models for other statistical institutes. Models are presented using the UML notation.

¹ Prepared by Georges Pongas (georges.pongas@cec.eu.int) and François Vernadat (francois.vernadat@cec.eu.int).

4. The paper presents some modelling elements towards the definition of statistical business objects in the domain of official statistics. This work is a conceptual framework derived from the analysis of 12 major statistical production systems of Eurostat. This model can be used for further refinement in the definition of statistical business objects and as a guideline in requirements analysis phases in the context of statistical system development.

5. Specifically for Eurostat, it is important to follow a uniform model that qualifies the states of the data sets, driven by a common meta-data model as defined in the paper, in order to allow external systems to retrieve information about the state of the data life cycle. Since little implementation work has been done so far on that topic, it is still possible to propose a software architecture that supports such a common life cycle meta-data model for Eurostat production systems. In any case, the system that deals with data life cycle monitoring must be fed by the production systems. The life cycle information supplied by the production systems must be independent from their internal data formats.

6. Future work at Eurostat concerns the development of an operational CVD Manager based on the meta-model presented here to record and control data flows evolution within Eurostat.
