

Distr.
GENERAL

CES/AC.71/2005/16 (Summary)
2 February 2005

Original: ENGLISH

**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 (ii): Development strategies for statistical information systems

**STATISTICAL META-INFORMATION SYSTEM AND DATA WAREHOUSE
AS MAIN SUBSYSTEMS OF SIS - CURRENT SITUATION AND PLANS**

Supporting Paper

Submitted by the Central Statistical Office of Poland¹

1. The Statistical Information System (SIS) is a global system, which will cover almost all statistical data processing systems in the Central Statistical Office of Poland (CSO). The statistical meta-information system and Data Warehouse are two main subsystems, which have been built in last 3 years.
2. One of the main goals of SIS was to integrate existing statistical distributed databases – to ensure that services offered by official statistics will be continuously available by improving access to the information already gathered.
3. The Data Warehouse facilitates performing analysis and secondary surveys based on distributed data sources through dissemination of the results of different reports from one place, where all necessary data are stored. Data, micro as well as macro, are available for statisticians in on-line mode. Data are secured from damage and from unauthorized access.
4. The Data Warehouse is metadata driven. It means that stored data are managed by metadata and every change of the data structure must be driven by previous metadata changes. The building of SIS was started from metadata integration (so-called subject matter integration) and the creation of Data Warehouse based on meta-information system (so-called physical integration).
5. Subject matter integration means the unique definition of terms and classifications used for documentation of statistical surveys, while physical integration refers to the internal consistency of stored data. The Data Warehouse is built in such a way that the same results can be derived in different ways – for instance based on individual (micro) data or based on aggregated (macro) data.

¹ Prepared by Krzysztof Kurkowski (k.kurkowski@stat.gov.pl) and Stanislaw Sieluzycycki (s.sieluzycycki@stat.gov.pl).

6. The metainformation system is the first system, where statisticians will be involved directly in the system. They will maintain a glossary of terms, classifications and other metainformation elements in on-line mode. The Data Warehouse will be maintained by IT people.
7. The metainformation system can be treated as an umbrella, which covers all other systems i.e. any metainformation used by other systems must be taken from one place only – the metadatabase. In turn, Data Warehouse is a core system where data are stored. There are, however, other subsystems, which will be included into SIS. These systems can be treated as front and back systems from the Data Warehouse point of view. Front systems are: designing and managing of surveys (including the survey documentation system); electronic forms systems for data feeding, editing, verification; and system of use of data from administration sources. Back systems are: reporting, analytical systems, data dissemination system (portal, content management systems (CMS), public data warehouse, including geographic information systems (GIS)).
8. The design of survey frames and survey documentation systems will be based on the metainformation system. Statisticians will have the possibility of creating frames using necessary terms as well as all information contained in the metadatabase. They will introduce survey documentation including creation of checking conditions, which have to be verified during data entry. Generally, such conditions are the most difficult elements of the system, as they require elaboration of internal language automatically interpreted and executed during the following stages of data processing.
9. Electronic forms system is based on the main assumption that internet systems will cover most statistical surveys in the near future. Although the trend is obvious, many implementation problems exist and, as can be seen in other countries, introducing such systems requires time. Electronic forms systems will derive information from survey frames systems as well as from the metainformation system. Security problems are also important as individual data are transferred directly through the internet. The CSO is currently implementing the first system (DG1), which covers economic data and where companies will send directly their data through the internet.
10. The use of administrative sources requires first of all many bilateral agreements between the CSO and other agencies. These include legal aspects, but also technical solutions. The use of the XML standard is a typical working solution.
11. With regard to back systems, the use of a reporting and analytical system is common. Generally, such systems depend on the relational database management system (RDBMS) used. The CSO Data Warehouse is built based on Oracle, so statisticians use the tools from this company. Oracle Discoverer allows for the on-line creation of reports directly by statisticians. Knowledge of the structured query language (SQL) is not required. Oracle Reports is used in more advanced cases where some programming is required (done by IT specialists). Reported data can be exported to Excel or comma-separated values (CSV) format also. By analytical tools we mean mainly the SAS statistical analysis package. Connection of Oracle with SAS is not used yet, but technically it will be possible soon.
12. Last but not least are data dissemination systems. Portal is the most important of these, as it shows not only documents managed by CMS, but also data and metadata taken directly from the Public Data Warehouse and metainformation system. The CSO is now preparing tender for the creation of a new portal, which should be built according to the newest technology and using the experience and best practices from other countries.
