Distr.
GENERAL

CES/AC.71/2003/12 (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 III: Efficient management of increasing technical complexity

TOWARDS A CORPORATE INTEGRATED INFORMATION SYSTEM ARCHITECTURE: THE CASE OF EUROSTAT

Invited paper

Submitted by Eurostat[1]

Summary

1. The paper presents the vision being deployed at the Statistical Office of the European Commission (Eurostat) about a rationalized IT infrastructure for integrated operations of its various statistical production systems. The new architecture being implemented isolates physical data from applications and users, uses database federation mechanisms, strongly relies on the use of metadata about storage systems, application systems and data life-cycles, emphasizes the use of thematic and support servers and uses a message-oriented middleware as its backbone for data exchange. Portal technology will provide the unique gateway both for internal and external users to have public or restricted access to information produced by over 130 statistical production systems working in the back-office. Architectural principles and solution are discussed.

2. Eurostat currently runs and manages about 130 software systems and applications. These systems act both on data, metadata and nomenclature items originating from the Member States, the candidate Member States and some other organizations (OECD for instance). The ultimate goal of Eurostat is to collect, assemble, manipulate and disseminate these data so that they get added value to be presented to a wide range of end-users (e.g., other Directorate Generals of the European Commission, European Central Bank, government agencies, banks, enterprises, politicians, press agencies, researchers and citizens).

3. A major problem is the independent and unstructured evolution of these systems and applications, and the many ways data are dealt with. This results in high maintenance costs and on-going ad hoc developments and practices with a risk of system "fossilization", i.e. rigid, stand-alone legacy systems. Another problem is the lack of synergy among these systems due to poor interoperability, no real integration and weak data mutualization.

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

4. The paper presents the IT solution given to these problems at Eurostat in the form of a new architectural framework for information systems. Ideas presented can easily be adapted for other similar intensive data processing environments.

5. Expected benefits of such a new architecture for an organization like Eurostat include:

- Modularity: The architecture is made of modular components (services, data stores, messages, data formats…) implemented on servers. This provides a high degree of independence from technology and allows easy modification or replacement of any component with minimum impact on the rest of the architecture.
- Flexible integration: Due to its modularity principle, the architecture caters for high integration of the Application System and Storage System without creating a huge monolithic information system, i.e. a networked system in which production system autonomy is preserved and which does not fall apart when one of its components evolves or changes.
- Transparency of service and data location: Thanks to the use of meta information and normalised access to data in data stores as well as to functional operations on application servers, the users or the applications do not have to know where data and services are located and which access methods are required to use them.
- Scalability: Because of its modular structure, the architecture is scalable both in terms of new services that can be added than in terms of data stores that can be accessed.
- Traceability: Thanks to its CVD Manager and Meta-Data Manager, the architecture can ensure a high level of traceability both regarding the life cycle of data sets processed and flow of processing made in application environments.
- Independence to development languages and systems: The architecture must rely on state-of-the-art standards of the moment and is intended for long-term solutions.
- Reduction of production and maintenance costs: Offering services on a shared basis to the many applications will lead to significant reduction of development and maintenance costs of statistical applications (i.e. the same update is made once).
- Major step forward to building a corporate data repository and management: By federating the various data stores used at Eurostat and managing the four environments (collection, production, reference and dissemination) under the same umbrella, the architecture gives Eurostat a way for better exploitation of the whole set of data put under its control and a means to increase its quality of services.

6. Key success factors identified include:

- realistic planning and design (in terms of time-frame, labour, budget and risks);
- strong and continuous support of top management;
- incremental implementation and deployment;
- continuity of service at the user level (i.e. current operations should not be stopped or disturbed);
- limited change of user operational habits, unless requested by the users themselves;
- good communication and explanation of what's going on (at all levels of the organization).
