

Distr.
GENERAL

CES/AC.71/2003/24 (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 I: Measures for the improvement of quality at the IT management level

FUTURE INFORMATION ARCHITECTURE

Supporting paper

Submitted by Central Statistics Office, Ireland¹

Summary

IT Strategic Objectives of the CSO

1. In the CSO Statement of Strategy 2001 to 2003 (available from the CSO web site), the CSO has set out its Information Technology Strategy objectives. The main outstanding objectives of this Strategy are:
 - to migrate and reengineer its existing DEC Alpha-based legacy systems to a client server environment
 - to implement the reengineered systems within the parameters of the CSO Corporate Data Model and its associated input, clean unit record, aggregate and disseminate databases and metadatabases and produce the applications necessary to support this
 - to consolidate CSO's approximate 100 surveys into approximately six business process classes
 - to build a common user application layer with functions that are common to all these classes and to build user application layers for additional functions specific to each class
 - to allow for web enablement in the receipt and dissemination of information
 - to interface the solutions proposed with some current client server systems and Sybase databases, such as the Central Business Register (CBR) and the Classification and Related Standards System (CARS).

These objectives together with migrating its existing data holdings to the new environment must be accomplished by the end of 2005.

2. The Statement of Strategy states that:
“...As mainframe systems are migrated to the new client-server environment, a generic approach to system re-engineering will be adopted. The common functionality required over different survey

¹ Prepared by Tom Keane (Tom.Keane@cso.ie).

systems will be created in a core template and then this will be copied and modified for each survey-specific processing system...”

and

“...The use of e-technology needs to be extended in both the collection and dissemination of statistics...”

Future Information Architecture

3. The future information architecture will encompass the applications, data and technology architectures required to realise the strategic information technology vision of the CSO. The key elements of this Future Information Architecture are summarised below.

4. The primary information architecture model required by the CSO is a data warehouse, front and back-ended with multi channel applications allowing the collection and dissemination of data from and to a variety of electronic sources as well as the more traditional channels such as paper.

5. Between the multi-channel layers will lie suites of applications specifically for the “management” of all the surveys. The term management includes the manipulation and editing of data as it evolves from individual survey responses and administrative sources, to statistical results, capable of being distributed over many electronic channels.

6. Underpinning this architecture will be a suite of productivity tools to enable the smooth running of the CSO.

7. Corporate data management requirements will be catered for in the new environment as all proposed systems will be developed within the parameters of the Corporate Data Model. This is outlined in Section III.2.

8. A scheme of the overall architecture is provided in the full version of the paper (English only).

Future CSO Data Management Framework and Overview of CSO Surveys

9. **CSO Development History:** The CSO conducts approximately 100 surveys covering both businesses and households. The processing of these surveys has been computerised independently in a DEC Alpha VMS environment using one of two approaches, viz.

- All systems development work done by local staff such as in the Agriculture and Foreign Trade Divisions (typically in SAS www.sas.com and TPL www.qqqsoft.com, etc.)
- Some or all systems development work done by the central System Development Section (SDS) (typically in Cobol).

Local staff also conduct further analysis on clean data files from systems that have been developed by SDS.

10. **Proposed Approach to Data Management under the Data Management System:** Systems will be re-engineered within the parameters of the Corporate Data Model and its associated input, clean unit record, aggregate and disseminate databases outlined below. In addition the metadata necessary to identify and explain all the statistical data in the four databases will be managed in metadatabases.

11. **Input Database:** This database (or group of databases) is where data will be initially loaded after data capture (including EDI or Administrative data), edited and updated.

12. **Clean Unit Record Database:** This database (or group of databases) contains a snapshot of the corresponding Input database tables after initial processing has been carried out. The Clean Unit Record database tables are static and date stamped for easy subsequent identification.

13. **Aggregate Database:** Initially the Aggregate database (or group of databases) tables will contain the first results from the aggregation (or tabulation) of the data in the Clean Unit Record database. Further

processing, e.g macro-level quality control and/or seasonal adjustment, will often need to be carried out on Aggregate database data. Thus it will be a live database. The Disseminate database will be populated with snapshots from the Aggregate database so there will be no further processing of data once it leaves the Aggregate database.

14. **Disseminate Database:** The Disseminate database will contain results data for all available periods. All Aggregate database tables are not necessarily brought into the Disseminate database. Some Aggregate database data may have been for internal analysis or checking purposes only. For audit purposes a date stamping system has to be applied to data in the Dissemination database tables. All data in the disseminate database tables is static.

15. **Metadatabase:** The metadatabase (or group of databases) will contain all system metadata and the metadata necessary to identify and explain all the statistical data in the databases. It will hold items such as descriptions of statistical variables, names of responsible experts, legal basis of surveys, etc. In some instances it will hold references or links to the relevant documentation.

16. **Other File Stores:** In addition to the four databases and the metadatabase above two file stores will be maintained to contain (i) all data transmitted to the CSO in the form of flat files, emails, etc.; and (ii) all unstructured disseminated data such as HTML or PDF reports, etc.

17. **Flow of Data through the Statistical Processing System:** A figure outlining the flow of data through the statistical processing cycle in the new data management framework is provided in the full version of the paper (English only).

- - - - -