UNITED
NATIONS

E

Economic and Social
Council

Distr.
GENERAL

CES/AC.71/1999/20
4 February 1999

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting on the Management of Statistical Information Technology
(Geneva, Switzerland, 15-17 February 1999)

Topic (iii): Integration of statistical activities at the national and international levels, including data modelling strategies and standards needed for statistical data integration

OECD DATABASES FOR STORING STATISTICAL DATA, METADATA AND CALCULATION METHODS

Submitted by OECD[1]

I. INTRODUCTION

1. The OECD has just terminated the migration of its main macroeconomic databases from its mainframe computer to a Client-Server system. The macroeconomic databases covered in the project are:

- *Annual National Accounts* (ANA): OECD standardised national accounts following the Standard of National Accounts;

- *Quarterly National Accounts* (QNA): national accounts data from National Statistical Offices following Member Countries National Accounts presentations;

- *Main Economic Indicators* (MEI): short-term economic statistics for Member Countries and some non-members. Contains data on production, prices, labour market, financial indicators, balance of payments, some national accounts items, business surveys results, etc.

[1] Prepared by Gérard Salou.

GE.99-

2. This paper describes the migration project in terms of objectives, strategy, achievements and organisation.

II. PROJECT OBJECTIVES

3. The project had the following objectives:

- **Improve the statistical service to users of OECD macroeconomic data**: One of the missions of OECD statisticians is provide a data service to the rest of the Organisation, to governments of OECD Member Countries and to the public in general through its publications. Macroeconomic data are a key element in all activities of the Organisation, in particular for economic analysis and forecasting. This project was aiming at improving the particular aspects of the data service that are data accessibility, data visibility and data documentation in general.

- **Improve the efficiency of statistical production**: It is a general objective for the management of the organisation to search for efficiency gains. Statistical production at the OECD is an area where technology and organisation have only evolved marginally in the last 20 years while the technology available for data management has made enormous progress. This project had the objective of making statistical production more efficient by making data management more structured so that less human intervention is needed.

- **Prepare the future**: A particular objective for the near future is to ensure that systems are Year 2000 compliant. A longer-term objective is to make sure that the OECD data systems will take advantage of future developments in technology. If this is successful then the first two objectives will be reinforced.

III. PROJECT STRATEGY

4. The three main elements of the strategy for reaching the objectives of the project were to:

- integrate data, metadata and calculation methods together as much as possible;
- structure data, metadata and calculation methods; and,
- use an IT environment and IT tools that are *de facto* market standards.

This section of the paper describes the rationale of the strategy.

III.1. Integration of data, metadata and calculation methods

5. Descriptive metadata, which include, for example, items like definition, coverage, or collection methods, are essential to locate, understand and efficiently use the related data. Production metadata, which include, for example, information about series published, technical information about the source, are essential for making the related operations more efficient. Some metadata can be descriptive as well as used for production. For example a calculation method can be used by the computer and also describe the related data.

6. Integrating calculations in the database also makes the data management more efficient. The nature of the OECD statistical work involves systematic operations on time series. Examples are rebasing to a common base year,

applying seasonal factors or linking indices to create long historical series. Integrating calculation methods in the database makes possible to standardise and automatise those operations. This also minimises the risk of errors in calculation coding. In addition, having standard methods makes the coding of calculations extremely simple and reduces the need for documentation. Therefore, it makes calculation methods easier to understand for the end user.

7. To take advantage of a database system to integrate data, metadata and calculation methods it is necessary to structure them. This is described in the following sections.

III.2. Structure data

8. The data model is multidimensional and aimed at having time series as main data object. The dimensions are the following:

- Country;
- Subject: is organised hierarchically to facilitate navigation and organise accounting data; it is itself multidimensional for national accounts data;
- Statistical measure: identifies the way indicators are expressed, for example seasonally adjusted, constant prices, in US Dollars, etc. or any statistically valid combination of them.
- Version: identifies successive generations of data corresponding to the same country, subject, and statistical measure.
- Frequency, to define a time series, and finally;
- Period in frequency.

9. The data structure is an essential element of the organisation of metadata and calculations as described below.

III.3 Structure metadata

10. Descriptive metadata is structured in such a way that standard metadata items are available to users for understanding and facilitating the identification of data. It follows the list of standard metadata items promoted by the OECD, see http://www.oecd.org/std/metagen.htm. Those items are available at various levels of the multidimensional structure as well as at various levels of the hierarchical dimensions. Descriptive Metadata identify the successive versions of indicators. Some items are available at the data point level as footnotes, data point qualifiers and information about the updating.

11. Tools have been developed to automatically produce pages of documentation on data series and to produce “Sources and Definitions” publications.

III.4. Structure calculations

12. Calculations made on macroeconomic data have been analysed and the resulting basic calculations classified. The most important types are listed below:

- Derivations, for going from one statistical measure to another of the same indicator; for example deriving an index in a common OECD base from the national base; of course, there is a large number of possible derivations with many standard methods;

- Links, for building long time series with successive versions of the same indicator; a number of standard methods are available, for example adjusting the series on the first common period;

- Frequency conversions, with standard methods to go from one frequency to another;

- Aggregation, for calculating an element of an account given the information contained in the hierarchical structure; several standard methods are available;

- Geographical aggregation, for calculating an indicator for a geographical zone;

- Free formula, for non-classified calculations.

Standard methods include rules of propagation of data points qualifiers.

13. It is also possible to do calculations outside the system for very complex calculations that do not need to be fully automatic. This makes necessary to be able to interface the database system with a wide range of user tools. This is facilitated by the next element of the strategy described below.

III.5. Use of de facto market standards as IT environment and tools

14. For hardware and software the strategy is to use de facto standards of the IT market. This particular project follows the overall strategy of the OECD in the area of IT, which is to minimise the number of platforms and operating systems.

15. The rationale for this choice was to minimise costs and to maximise the number of products and services compatible with the architecture. In addition, working with the standards makes it easier to rapidly benefit from progress in technology. The increase in PC power is an example of benefit which can be obtained at low cost by using the market standard for work stations.

IV. TECHNICAL ENVIRONMENT.

16. The OECD macroeconomic databases have been developed in accordance with the OECD overall IT standards for databases, operating systems and computing languages. The systems used are:

- Hardware: networked PCs, in Client/Server configuration;
- Operating system: Windows NT 4.
- Relational database management system and client/server application layer: Microsoft SQL Server 6.5; It is being migrated to SQL Server 7 (February 1999).
- Programming language for main client: Microsoft Visual Basic 5.
- Version management: Visual source safe.

17. The main user tools used for analysis at the OECD are Fame and Microsoft Access and Excel.

Figure 1 shows the general architecture of the system.

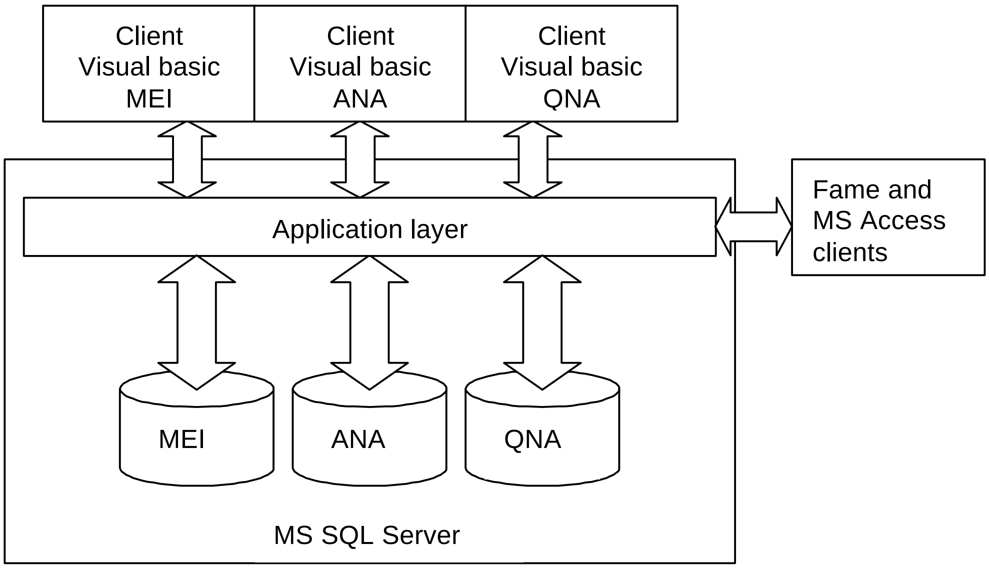

Figure 1. *General architecture of the OECD macroeconomic systems*

V. FUNCTIONALITY DEVELOPED

18. The main client to the database has been developed using Visual Basic version 5. The functionality is given by a combination of the database, database application and client. The following is a summary:

- Navigation in the database: selecting data in the multidimensional space;
- Data and metadata entry, graphing, reporting, exporting;
- Time series management: calculations, creation, deletion, version management;
- Database global objects management: countries, subject hierarchy, units, measures, user profiles.

19. A particular client to the database is the software package Fame. It is the main tool for analytical and forecasting work at the OECD. It provides the complex functions that the database system cannot handle and automatise.

VI. DEVELOPMENT METHOD AND PROJECT PLANNING

20. The project was based on an adaptive/iterative implementation approach with prototyping and direct user involvement. Statisticians as well as users of data in the Organisation supplied initial specifications. Then, the overall development method was to use rapid application development techniques, and in particular, prototyping, to ensure that a maximum of user requirements gets satisfied. A particular management problem was created by the fact that the statisticians responsible for each of the three main data sets had expressed different needs in some areas. A steering group, composed of managers in the Statistical Directorate was responsible for solving the resulting conflicts and for monitoring the project.

21. The project started in April 1997. The most important milestones in the

project were the following:

1997

- April: official start of project;
- August: end of database design work;
- December: presentation of version 0.1 to statisticians and end-users.

1998

- February-August: successive versions tested by statisticians;
- August: end of new development, beginning of data/metadata/calculations migration phase;
- November-December: parallel work;
- December: end of production on mainframe system, archiving of mainframe databases;

1999

- January: general deployment of the new applications.

22. The project team was composed of about ten persons, including consultants hired for specialised tasks and for advising on the optimal use of the technology chosen.

VII. CONCLUSION

23. The new databases were put in production in January 1999, the applications deployed and the mainframe stopped. The migration of data, metadata and calculations has already permitted to rationalise and improve the calculation of derived indicators. The new systems are clearly faster for data input and verifications, and also more flexible for output and preparation of publications. However, work continues to consolidate the system.