

Building OECD's New Statistical Information System

Prepared by Lee Samuelson (lee.samuelson@oecd.org) and Lars Thygesen (lars.thygesen@oecd.org)

1. Introduction

The gathering and harmonisation of international statistical data in a multidisciplinary environment are key to international comparative analysis and policy work. The availability of timely, accurate statistical information enables OECD committees, officials in Member countries and the OECD Secretariat to address a wide range of issues in today's rapidly-evolving global economic and social landscape

The OECD is developing a modern Statistical Information System to support its statistical activities. The objectives are to improve the *efficiency* of data and metadata collection, validation, processing, storage and dissemination; improve *quality*, eliminating errors and incoherencies and shortening statistical publication cycles, and enhance the *accessibility and visibility* of the Organisation's statistical outputs.

In order to achieve this, the Organisation is taking full advantage of advances in information technologies and standards (e.g., OLAP, XML, Web Services, SDMX-ML, GESMES/TS).

The structure and ideas behind the new Statistical Information System are inspired by best practices in the international community of official statistics.

The overall architecture of the Statistical Information System consists of three layers:

- a **production layer** for collection, validation, processing and management of statistical data and metadata
- a **storage layer** where validated statistics and related metadata are stored
- a **dissemination layer** for producing statistical publications and online/offline interactive statistical products

The three layers (or pillars in the architecture model, see Figure 1) are supported by a workflow system which automates statistical and publication processes wherever possible, and tracks the steps involved.

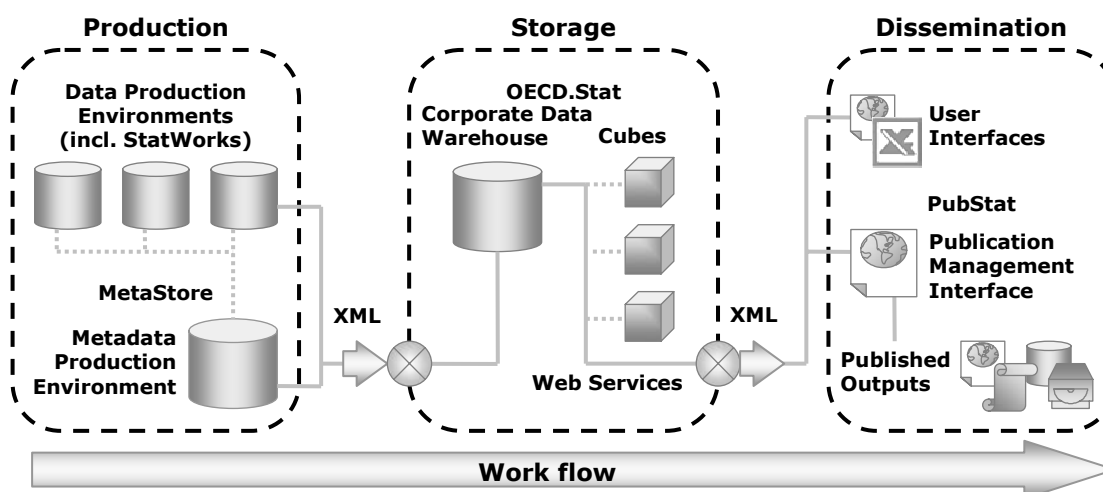


Figure 1. Architecture of the new Statistical Information System

The use of Web Services and XML permits "loose coupling" of the System's components. Thus a change in one component or its inputs and outputs can be made without necessitating changes in the others.

2. Components of the New Statistical Information System

The five independent but inter-operating components of the new Statistical Information System are described below.

a. OECD.Stat – the central repository for validated statistical data and metadata

OECD.Stat is the central repository ("warehouse") where validated statistics and related metadata are stored. It will, in due course, be the sole and coherent source of statistical data and related metadata for the Organisation's statistical publication and electronic dissemination processes.

OECD.Stat enables the Organisation's analysts and statisticians to easily locate needed data from a single online source, rather than having to navigate multiple databases and data formats, each with its own query/manipulation systems. And the access to systematic metadata in OECD.Stat helps ensure appropriate selection and use of statistical information.

OECD.Stat has been designed to preserve the decentralised nature of OECD directorates' statistical activities, while making their data and metadata part of a coherent corporate system. OECD.Stat is being progressively populated from the nearly one hundred time-series and cross-section production databases managed by directorates across the House.

Metadata can be attached at any level of an OECD.Stat dataset: dataset-level (e.g. national accounts), dimension-level (e.g., "Variable means..."), dimension member-level (e.g., "GDP means..."), and data-level for any lower level (e.g., series level, observation level, etc.).

The present user interfaces of OECD interactive statistical databases are not sufficiently easy to use for outside users or users who are not already familiar with the structures of the parent production databases. The OECD is seeking to remedy these problems by providing access to OECD.Stat via a state-of-the-art presentation tool for the electronic dissemination.

b. StatWorks – modernising the production database environment

OECD statistical database applications have been developed using many different software platforms (Oracle Express, SQL Server, Excel, Paradox, Access, FAME, etc.). This is a reflection of the decentralised nature of the Organisation's statistical activities. StatWorks has been developed to provide a common hosting environment for production databases, based on the OECD standard database platform, MS-SQL Server. The objectives are to strengthen standardisation of design and minimise the number of tools used in statistical activities, and correspondingly reduce training and support requirements. OECD directorates are being encouraged to migrate production databases hosted in non-standard environments to StatWorks. However, there is no intention of forcing migration of all the present production environments to StatWorks, as some of the production systems provide very sophisticated support for the specific statistical area, which could only be migrated through a major development effort.

StatWorks includes a toolkit for managing statistical data, which minimises the need for developing and supporting database-specific programs. This toolkit includes facilities for:

- initial data migration
- database administration (managing user access rights, defining dimensions, adding a country, etc.)
- security management
- data collection via interactive questionnaires
- data validation
- data manipulation (e.g., calculation of regional totals)
- dynamic links to MetaStore
- export of datasets to OECD.Stat

c. MetaStore – coherent management of metadata

OECD's corporate metadata facility, MetaStore, is designed to improve the efficiency of metadata preparation, storage, access, management and dissemination for the Organisation's statistical products. MetaStore provides dataset managers with a common set of tools for managing metadata, and supports adherence to common standards for statistical metadata across the House. MetaStore addresses problems of fragmented metadata located in numerous databases and text files maintained by different Directorates, avoiding duplication of effort in metadata preparation, gaps in metadata availability, and inconsistent metadata across databases.

MetaStore supports a set of *metadata management principles* which apply to metadata describing validated statistical data to be shared or disseminated, whether internally or externally. These principles imply that all statistical data must have appropriate metadata, covering a broad scope of user needs and that statistical metadata must be consistent across subject matter areas. A set of guidelines help database managers to accomplish these goals. A common structure of the metadata, defining common metadata items and arranging the metadata under headings and subheadings like unit of measurement, data source, etc., is included. The structure will increase the usefulness and ease the exchange and sharing of the metadata, and it is closely linked to international standardisation work under the SDMX initiative mentioned below.

MetaStore provides a database manager with:

- a storage area for statistical metadata
- an interface for managing production metadata
- facilities for enriching current metadata content through external links to standard classifications, glossary terms, SDDS concepts, etc.
- export of metadata to OECD.Stat

Like OECD.Stat, MetaStore can accommodate any kind of metadata related to any level of detail of the corresponding statistical data. For example, metadata may pertain to a database in general (i.e., purpose of the database, database manager, most recent update, next update, dimension members, methodology used in compiling statistics, seasonal adjustment methods, publications derived from the database, etc.), or be specific to the data themselves (i.e., breaks in series, missing observations, estimates, revised figures, etc.).

d. PubStat – modernising statistical dissemination processes

Technology and publication standards for the OECD statistical products have been defined, enabling the development of modernised tools and processes for producing traditional statistical publications and interactive data products. This is made feasible by the existence of a single source for validated statistical data and metadata, OECD.Stat. Thus PubStat is developed with the objectives of:

- increasing the efficiency of statistical dissemination processes
- reducing the risk of human intervention and, thus, of mistakes
- reducing time-to-publish
- giving the Organisation's statistical publications and electronic products a common "look and feel"
- reducing the number of different software tools and corresponding support effort involved, and
- minimising the time that statisticians spend dealing with dissemination and formatting issues.

PubStat contains, for a given publication, information on the structure of the tables, the statistical data and metadata to be extracted from OECD.Stat, where these data are to appear in a statistical table, the “publication metadata” containing headings and labels, and style sheets implementing the graphical presentation of the tables.

PubStat provides the database manager with:

- an interface for managing publication contents definition
- a facility for storing these instructions for later re-use
- an interface for generating an XML output file combining statistical data and metadata with publication layout instructions

PubStat, in conjunction with OECD.Stat, will make it possible to streamline the production of the Organisation’s existing “horizontal” statistical publications, where statistics are drawn from several separately-managed datasets (e.g., “OECD in Figures”), and supports production of new ones (e.g., “OECD Factbook”).

e. Workflow – automating statistical and publications processes

A workflow system is being developed to automate the whole statistical and publication process wherever possible, and track the steps involved. This workflow system will be used to initiate and monitor the transmission of data from one layer to another – notably from the Production Layer (StatWorks, other production systems, MetaStore) to the Storage Layer (OECD.Stat), from the Storage Layer to the Dissemination Layer (PubStat), and from PubStat to the various channels used for dissemination of the Organisation’s statistical products: PDF files for traditional publications, specialised files for online dissemination (Internet, OLIS, SourceOECD), and specialised files for dissemination via CD-ROMs. Dataset managers can choose to receive notification by e-mail of the results of each process.

3. Impact on cooperation with Member countries and other international organisations

Closely related to the development of its new Statistical Information System, OECD is committed to making the best possible use of information technologies to facilitate the exchange of statistical data and metadata, and to sharing best practices with Member countries and other international organisations. Indeed, the problems which the OECD seeks to solve by building its new Statistical Information System seem to be general and exist in many statistical environments world wide. The system is modular and built using well-established standards, making exchange of components relatively easy. Therefore it is expected that the components would be reusable in national and international organisations (especially for managing decentralised statistical systems). Along this line, cooperation has already been initiated with UNESCO which will make use of some of the components mentioned above. OECD is also looking for best practices among peers in order to integrate such practices and as far as possible avoid duplication of work.

In addition, a number of software tools and standards for management and exchange of statistical information are being developed cooperatively with other international organisations. Following agreement to share responsibilities for collection of annual foreign trade data, the OECD and UN have additionally agreed to work jointly to establish a common system for managing annual foreign trade data, using an SQL-based data model designed by the UN Statistics Division (the *ComTrade* project). Development effort for the data collection, validation, processing and management software is being shared by the OECD and the UN. Responsibilities for collecting and validating data will also be shared, with data periodically replicated from one site to the other.

OECD is actively working with Eurostat to minimise duplication of data collection, and sharing data collected concerning short term economic indicators and structural business statistics. The OECD has already successfully implemented data sharing arrangements with other organisations on a smaller scale. OECD currently provides Index of Industrial Production statistics for OECD countries to the IMF in lieu of their direct collection of these data and receives all business tendency and consumer opinion survey data for European Union countries direct from the European Commission.

Most importantly, OECD is fully committed to the SDMX¹ initiative sponsored jointly with six other international organisations² and chaired by the OECD, aiming at finding common solutions for exchange and sharing of statistical data and metadata - and making them work.

The development and implementation of new data exchange and sharing standards is a common task of the OECD and its Member countries, and real progress can only be accomplished through partnership between these parties and other international organisations. With this aim, OECD has recently set up an international expert group -- the OECD Expert Group on Statistical and Metadata Exchange³ -- to foster dialogue with national and international partners on strategic issues related to the development and practical implementation of new technologies and new procedures for statistical data exchange, and to share best practices in this regard. The Experts Group began its work in April 2004.

These efforts to modernise and standardise both internal and external statistical systems and procedures are expected to have a positive impact on the way the OECD works with Member countries and other international organisations.

¹ Statistical Data and Metadata eXchange, see www.sdmx.org

² BIS, ECB, IMF, Eurostat, UN, and the World Bank

³ See www.oecd.org/std/research/exchangeexpertgroup