

Distr.
GENERAL

Working Paper No.10
26 April 2007

ENGLISH ONLY

**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**

Meeting on the Management of Statistical Information Systems (MSIS 2007)
(Geneva, 8-10 May 2007)

Topic (ii): Statistical information systems architecture

**IMPLEMENTING A STATISTICAL DATA WAREHOUSE:
THE CASE FOR A STRONG PARTNERSHIP**

Invited Paper

Prepared by

IMF: Cathy Wright, Statistics Department and Patrick Hinderdael, Technology and General Services
Department

OECD: Lars Thygesen, Statistics Directorate and Douglas Paterson, Information Technology and Network
Services Directorate

I. BACKGROUND

1. The OECD and the IMF are intensifying their cooperation in statistics, developing a partnership for sharing data as well as data warehouse technology. This development takes place on the basis of a long-standing cooperation of data exchange and standardisation between the two organisations.

OECD

2. The OECD established its statistical data warehouse, OECD.Stat, in 2004 to provide internal access to the institution's wealth of diverse data sources in a central environment. OECD.Stat is part of the OECD Statistical Information System, which covers the full life cycle of statistical data. It also encompasses a data production environment StatWorks, a metadata management system, MetaStore, and a data dissemination environment, which presents the data in the warehouse for different media. An important part of the dissemination environment is the web browser, allowing internal and external users to search and browse

across datasets, view data and export in multiple formats, including SDMX-ML. Direct links from analytical software tools such as FAME, Excel and SAS are being developed to facilitate use of the warehouse for data sharing.

3. In 2005, access to the data warehouse was extended to organisations participating in the OECD's extranet for government officials (OLIS), and selected parts of the data warehouse, i.e. the data that can be made freely available to the public according to OECD dissemination policy, were made publicly available on the OECD web site in 2006. It is planned in 2007-2008 to make OECD.Stat the unique platform for all dissemination, including the OECD commercial web site SourceOECD.

4. The data warehouse covers almost all statistics of the OECD, and work is ongoing to include the rest. The system includes recommendations and guidelines on quality, metadata, and governance¹ for the whole organisation. Functionality of the data warehouse and its web browser is also being continually improved according to users' needs.

IMF

5. The concept of developing an economic data warehouse was born in the winter of 2005 with the new Director of the Statistics Department, Rob Edwards, who had overseen the introduction and benefits of a data warehouse at the Australian Bureau of Statistics prior to coming to the IMF. Inspired by his experience, a small project got under way with a fairly modest agenda: 1) to gain a better understanding of data warehouses, especially those containing economic data, 2) to learn about data models and the role of metadata, 3) to begin to gather user requirements, and 4) to build a prototype warehouse. The early days of the project were enormously challenging. Researching the institutions that had introduced data warehouse and learning of their experiences was an especially important aspect of the work as no one at the IMF possessed first-hand knowledge or experience with such an environment.

6. With the benefit of a brief consultancy from Margaret Salmon of the ABS, a prototype data warehouse was developed. The prototype, while not operational, provided a platform from which one could demonstrate the features and potential benefits that a data warehouse could bring to the IMF. It was determined that certain features would be paramount to the success of the data warehouse at the IMF:

- Searching across databases for a particular country or concept, being able to return results to the questions "What data are available for Mexico?" or "What data are available for foreign direct investment in the WEO and IFS databases"?
- Introducing a metadata-driven environment that permits the use of structural and/or referential metadata to search for and elaborate data results.
- Building an environment that conforms to the emerging SDMX standards.
- Providing a foundation from which publications and electronic data products can be disseminated.

7. Following a mission to visit the OECD, ECB, BIS, and Eurostat in the spring of 2006, the OECD graciously offered to share its technology with the IMF and the IMF accepted. Upon further discussion by the directors of the Statistics and IT departments, work began in earnest in June 2006.

¹ www.oecd.org/statistics/qualityframework

II. THE BUSINESS PERSPECTIVE

8. The Statistics Directorate/Department of the OECD and the IMF are engaged in fundamentally the same business. Both aim to provide timely, high-quality, internationally comparable and understandable data to their internal and external clients. While the content of the databases and the production and dissemination environments differ to a greater or lesser extent, the underlying business processes, drivers, and outcomes are identical. Core tenants of our business environments are to: minimize the reporting burden on member countries, enhance coherence of data between countries over time and between domains, eliminate obstacles to data access, and communicate effectively with data provided to users.

9. As the data needs of the two organisations overlap in many domains, data sharing has been going on between them for a long period of time. Because of incompatibility of technical platforms, the data sharing was originally carried out in archaic forms, involving considerable resource spending and loss of timeliness. In 2006, an SDMX based mechanism was set up between the databases of the two organisations to allow for easy and quick exchange. While SDMX can be used to interlink databases of different designs, a more intensive data sharing could be conceived if database systems were also shared.

10. The roles of statistical departments of both organisations are much the same. They conduct certain basic statistical activities themselves and have a role of coordinating the whole spectrum of statistical activities across their organisations. They coordinate with their information technology department (OECD Information Technology and Network Services [ITN] and IMF Technical and General Services Department [TGS]) to design and deliver the necessary tools, as well as with their public facing bodies (OECD Publications and Communications Directorate [PAC] and IMF External Relations Department [EXR]) to implement data flows to external communities of users. Both organizations are among the seven international institutions sponsoring the Statistical Data and Metadata Exchange (SDMX) initiative.

11. Thus, it became apparent early in our discussions that a solid business case could be made for collaboration, first by setting the stage on the technology front but with a clear view to enhancing the data sharing across both organizations. This clearly has merit in its own right but becomes especially important during times of budget constraints faced by both organizations. The ultimate vision is to explore the feasibility of creating joint databases and products.

III. THE TECHNOLOGY PERSPECTIVE

12. Since both organizations have been involved in much the same business for many years, with regard to statistical information management, it is not surprising that they have developed IT systems in parallel that share many of the same requirements. With the advent of SDMX as an international standard along with a maturing of the underlying technology platforms, it became apparent that an opportunity existed to lower costs and shorten development time while maintaining a greater level of support and capabilities by close collaboration between the IMF and the OECD.

13. To this end it was determined that installing and using a software system that was already developed by the OECD could potentially lead to a faster ramp up time for the IMF's data warehouse endeavor. Moving forward, future development costs could also be lowered and software quality enhanced through shared development between the organizations, which will eventually lead to a stronger platform to integrate with various publication and analysis systems.

14. The key to leveraging this collaboration is the alignment of goals and standards between the organizations covering all of the following:

- **Governance Models** – both organizations see the data warehouse as an “Enterprise” system that touches almost all of the functional areas of each organization.
- **Business Processes** – at a high level both organizations have similar statistical production business processes; and the specific facets where they do differ do not present an insurmountable complexity so that they cannot be managed differently at each organization. For example, the OECD leaves data owners responsible for filtering, formatting, and loading data into the data warehouse whereas the IMF has left this responsibility to the IT and Statistical departments. This difference in policy does not negate the effectiveness or efficiencies gained in implementing the same system at each organization. Another example is that the OECD has organised data in a number of multidimensional datasets (cubes), each related to a specific domain and with a specific set of relevant dimensions; whereas the Fund has taken a different approach in terms of mapping all data sets to a common set of dimensions using its existing Catalog of Economic Time Series although it remains to be seen whether it will be able to maintain that approach as the scope of the data warehouse increases.
- **Technical Implementations and Processes** – It was found that both organizations use much the same development/integration platform and almost identical runtime environments including: MS Windows Server and Clients, MS SQL Server, MS .Net Framework, and MS Office suite of applications. XML and Web services are seen as the “glue” supporting development of modular applications. Utilization of the collaboration tools available in this framework, namely MS Team Foundation Server, has already proved to be an invaluable mechanism to perform joint development between the two continents. Team Foundation Server facilitates the sharing of the source code between the two organisations, with separate branches that can evolve in parallel and be re-synchronized when appropriate. It provides a rich collaborative environment (MS Sharepoint) that greatly enhances communication and sharing of information between the teams as well as providing the basis for an “industry quality” collaborative software development process. It also includes automated build and test facilities to ensure software quality.
- **Communications Model** - While the semi-informal arrangements between the technical teams in both organizations has served all well so far, it is expected that a more refined approach will be necessary when and if more partner organizations become involved. It is recommended that we attempt to not only align the code-base but also functional requirements, use of development resources, deployment schedules, and data resources to maximise the efficiency gains and so that any new partners have a clear understanding of their role and responsibilities while also maintaining the ability to manage overall expectations of the software system among partner organizations. At the same time, the collaborative model must allow each team the flexibility to respond rapidly and effectively to evolving internal functional requirements.

IV. THE OUTCOME TO DATE: RISKS, COSTS, AND BENEFITS

15. After many months of discussion and exchange of software, in December 2006 OECD staff from the business and IT directorates came to IMF to determine whether the full collaborative endeavor could work and, if so, how the arrangement would work in practice. Aside from technical matters, decisions would need to be taken about the exact nature of the collaboration, resource commitments, communications, timing, etc. These elements are being embodied in a Memorandum of Understanding (MOU) drafted by the OECD and IMF and under review by the respective legal departments. The signing of the MOU is expected to take place in early summer 2007.

16. To date IMF have implemented and deployed the majority of the functionality of the OECD.Stat system along with two limited versions of time series from the World Economic Outlook (WEO) and the International Financial Statistics (IFS) data management environments. While there is still much work to be

done to operationalize the rest of the system, we believe that the currently available features are already able to demonstrate the value of a data warehouse to Fund end users. The short-term plans call for completing the deployment and configuration of all the systems features along with continued mapping and loading of data determined to be most valuable to Fund users. This will probably also include REO (Regional Economic Outlook) data, data from external service providers (e.g., Haver, DataInsight, Bloomberg) and data from desk economists/area departments.

17. The OECD has been following closely the implementation work in the Fund, with a staff member on site for a month at the outset of the project and advice provided as required. Teleconferences to discuss progress are held on a weekly basis. Wish-lists from the Fund of future enhancements have been discussed and have proved to fit well with OECD's own needs. As mentioned above, the systems are already working in production mode at OECD with a large number of internal and external users. Feed back from those users as well as ideas from IMF lead to continued enhancement of the system and its components. In accordance with the MOU, OECD informs IMF about and discusses planned and realised advances.

18. Both parties are now looking forward to the formal decision by the IMF to put the system in production, and to the finalising and signing of the MOU. It should pave the way for enhancing the data sharing. In addition, it is expected that the two parties (and others who might wish to join in) maintain a joint development plan, where each party can contribute in domains of special interest, while keeping the coherence of the system.
