

Distr.
GENERAL

WP.35
9 April 2010

ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION
FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN
UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Meeting on the Management of Statistical Information Systems (MSIS 2010)
(Daejeon, Republic of Korea, 26-29 April 2010)

Topic (ii): Software sharing and shared maintenance

Sharing Statistical Software – an update on the OECD experience

Prepared by Trevor Fletcher, OECD

I. Introduction

1. The OECD has been sharing its main Statistical Information System (SIS) software components with a number of other International Organisations and National Statistical Agencies since 2007. The various SIS components manage the full life-cycle of statistical data and metadata processing from data capture through dissemination.
2. A report of this activity was presented at MSIS 2009 in Oslo (See attached link to the original paper: <http://www.unece.org/stats/documents/ece/ces/ge.50/2009/wp.7.e.pdf>). This paper provides an update on the status of OECD software sharing in the last 12 months. Requests to evaluate SIS have increased considerably during this period and the paper summarises the background on how the sharing of statistical software came about and explains the advantages and challenges of the process. The paper highlights the advantages of sharing software for saving costs and promoting standards, in particular SDMX. The paper looks at and the possible role of the Sharing Advisory Board in this activity, briefly describes the software components being shared and examines the issues OECD is facing in sharing software. The paper seeks input from MSIS as to how collaboration could best be managed in the future.

II. Background

3. The OECD began sharing the main components of its Statistical Information System software in 2007 when the IMF requested to collaborate on the development of the *OECD.Stat* statistical data warehouse project. This followed an earlier collaboration project on developing a Trade system with UNSD which proved to be successful and thus encouraged the OECD to embark on further joint ventures.
4. A Memorandum of Understanding (MoU) was drafted between the two organisations and it was agreed that future developments would be managed on a shared basis as a single project. The MoU covered the following aspects of the collaboration:

- Mutual interest of participating organisations
 - Responsibilities of the participants
 - Governance
 - Apportionment of costs
 - Intellectual property
5. There have subsequently been numerous expressions of interest from other International Organisations and National Statistical Agencies. Following the experience in with the IMF it was decided not to try and run these various exercises as a single project as this proved to be more complex than had been anticipated.
 6. Initially each requesting organisations had a specific MoU drafted customised to the particular collaboration. With the increase in requests from organisations to evaluate the OECD Statistical Information System software a generic MoU was proposed to streamline what had become something of a lengthy process involving the Legal services. The generic MoU covers the following aspects of collaborating on OECD SIS software:
 - Cooperation activities (scope including support cost recovery & sharing expertise and experience)
 - Intellectual property
 - Disclosure
 - Responsibility
 7. During 2009 the requests to use OECD Statistical Information System software, in particular the *OECD.Stat* data warehouse, increased sharply and is leading the OECD to re-think its policy on how software sharing can be best managed.

III. Advantages of software sharing

8. As noted in the 2009 paper, there are a number of very obvious reasons for organisations to share statistical software and this activity is growing considerably among International Organisations and National Statistical Agencies. These include:
 - Savings in time and money in re-using or adapting an existing technical solution to meet another organisations processing requirements.
 - Sharing new features developed by one of the collaborating parties by the other partner(s) at minimal cost.
 - The use of common shared systems can promote and enable the use of statistical standards, both in structure and content. As an example, SIS components widely use the Statistical Data and Metadata Exchange (SDMX) standard as inputs and outputs.
 - Using the same software platform for statistical data management makes the exchange of data and metadata easier between organisations and can facilitate the production of joint data collection or dissemination exercises between organisations.
 - Collaboration between organisations in this way sends a very positive message to stakeholders that we, as publically funded bodies, are making our best efforts to efficiently use resources by working together and not duplicating effort in re-inventing the wheel.

IV. The Sharing Advisory Board

9. The MSIS Sharing Advisory Board (SAB) was launched at the Eurostat IT Directors Group meeting in October 2009 with the specific goal of improving the level of co-operation and sharing of software and related solutions between statistical agencies and international institutes. The mandate for this work comes from the Conference of European Statisticians. See here for a link: <http://www1.unece.org/stat/platform/display/SAB/Home>

10. The goals of the SAB very much coincide with those of the OECD in the sharing of its SIS software and believes it should be further explored how the SAB can be used as a mechanism to further promote collaboration between organisations.

V. Promoting use of SIS to enable SDMX exchange

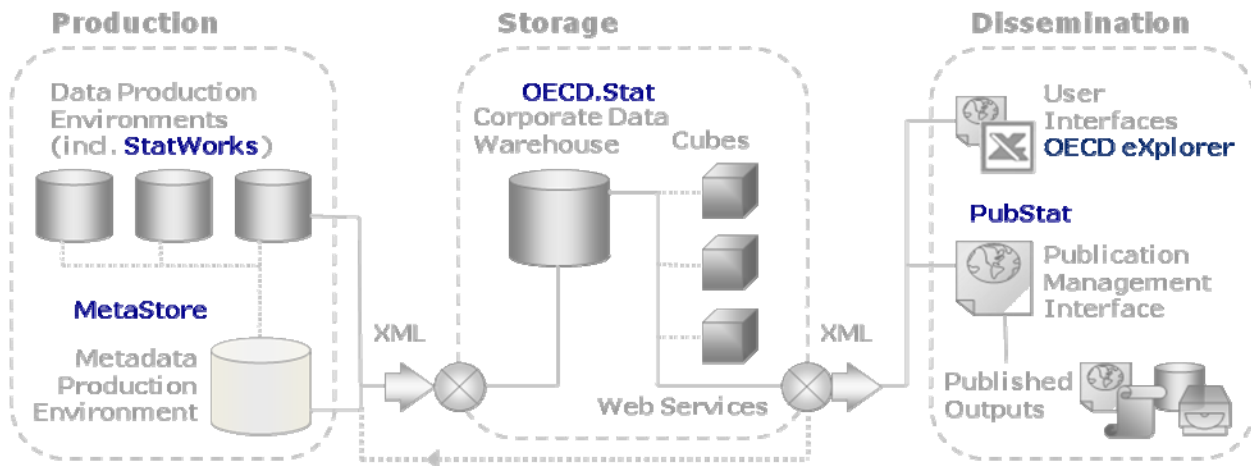
11. As already noted, sharing SIS software with other organisations is seen by the OECD as means for promoting standards for data exchange. This refers particularly to SDMX.
12. SDMX is the Statistics Data and Metadata eXchange standard. From the OECD standpoint SDMX provides a standard model for statistical data and metadata exchange between national agencies and international agencies, within national statistical systems and within organisations. SDMX is the preferred standard recognised by the UNSC and its main goals are to facilitate data and metadata exchange; to make efficient use of technologies and standards; to reduce the reporting burden for national agencies; and to enhance the availability of statistical data and metadata for users.
13. The approach being taken by the OECD with regard to SDMX is that any new statistical process or software development should incorporate SDMX as the data exchange mechanism and that existing systems should provide for SDMX both as inputs and outputs. This is leading to a steady increase in the use of SDMX within and outside the Organisation. These initiatives include:
- Existing data exchange mechanisms between the OECD and IMF based on CD-ROMs are currently being phased out and replaced with a data-warehouse-to-data-warehouse mechanism (Web Service) using SDMX.
 - Resellers of OECD data have been testing during the last 6 months a mechanism to receive data via the SDMX web service rather than by CD-ROM as at present as part of a planned adoption of SDMX as the standard format for receiving OECD data.
 - Tests are underway to replace the existing commercial CD-ROM dissemination platform (that uses a proprietary format) with SDMX and a generic reader that can be installed on any platform.
 - SDMX inputs and outputs are planned or have been implemented for the *OECD.Stat* Data Warehouse and the *Statworks* production software.
 - Tests are underway to use SDMX in a Trade In Services pilot data collection exercise with a number of National Statistical Agencies.
 - Tests are underway to use SDMX in joint Principal Global Indicators (PGI) data hub together with IMF, World Bank, Eurostat, ECB, BIS and OECD.
 - SDMX is being used as a generic input to the OECD *eXplorer* data visualisation application being developed for the OECD by the NCVA research institute in Sweden.
14. Other International Organisations are developing similar projects to enhance their capacities to receive, disseminate and share data in SDMX. However, as recognised in the March 2010 joint OECD/UNECE SDMX Experts meeting in Geneva, there are still several challenges to making SDMX the global standard implemented worldwide. These include:
- Although considerable activity was noted since the last Expert Group meeting held in 2008 there is still a lack of progress in national statistical organisations publishing data in SDMX. A goal of the meeting was to determine the reason for this and to agree a plan for moving forward.

- There was general agreement that the reason for the slow implementation of SDMX dissemination by National Statistical Agencies is content rather than technology: there are insufficient Data Structure Definitions (DSDs) to allow organisations to map their data to SDMX and the process to agree DSDs between organisations can be slow. Work is required to accelerate this process among the various domain experts – it is not a task for the SDMX sponsors.
- National statistical organisations requested that they be involved more in the governance and decision-making processes of SDMX. Their formal involvement is required in taking the standard forward. This could also increase the resource pool and availability of funds for future SDMX development work.
- Senior management in national statistical organisations need to be convinced of the business case for SDMX in order to promote its use within their own organisations. It was widely noted that this was a key factor in the slow rate of implementation. This would lead to a change in perception on the use of SDMX within national statistical organisations and an emphasis in shift from data exchange to the benefits of using the SDMX model in the end-to-end business process. UNECE proposed organising a seminar for Chief Statisticians at the next available opportunity.
- National statistical organisations requested that international organisations harmonised better their own requirements for DSDs.
- There was a general request that existing DSDs defined across all the sponsor group organisations and national statistical organisations should be centralised in a registry and this made publicly available.
- Following the presentation of SDMX and the semantic web it was proposed that a version of SDMX for linked data (SDMX-RDF) be developed.
- Participants should make others aware of SDMX tools by making them available in the online database, which will shortly be launched by the Bank for International Settlements, and linked to the SDMX web site.

15. It is hoped these actions will contribute towards the full implementation of SDMX at national and international levels. The extent of the benefits coming from the implementation of SDMX will be proportional to the number of organisations using these standards to exchange data.

VI. Software being shared by the OECD

16. Up to 2009 the main components of the OECD Statistical Information System being shared were those that manage statistical data and metadata Production, Storage and Dissemination processes (*Statworks*, *Metastore* and *OECD.Stat*). The SIS package has recently been enhanced by the inclusion of the OECD ‘*eXplorer*’ data visualization software. The following diagram demonstrates how each component fits into the overall OECD Statistical Information System architecture:



17. The *Statworks* software provides a common, SQL-based repository for statistical quantitative and qualitative data together with a set of tools for data management that covers Data Importing, Validation, Calculations, Querying and Exporting. The application is designed to be fully integrated with other components in the OECD's Statistical Information System (SIS)
18. *Metastore* is a general toolkit with a web-based interface for accessing and managing structural and reference metadata for statistics. Structural Metadata describes the multi-dimensional structure of a dataset and includes codes and names of dimensions and corresponding dimension members. Reference Metadata consists of textual information documenting characteristics of the data within a dataset and typically includes information concerning the collection, manipulation, purpose and quality of data at different levels of a dataset's multi-dimensional structure.
19. *OECD.Stat* is the corporate statistics data-sharing and dissemination environment of the OECD. Its data warehouse provides a harmonised and platform-independent access to shared or publishable data from all sectors of OECD activity and from other International Organisations. It is fully integrated with the other components of the OECD's Statistical Information System for managing production data and metadata, analytical software packages, and for the dissemination of Statistics via the Web, Paper and Electronic Data Products..
20. *OECD eExplorer* is a web-based interface for exploring and analysing statistics. This interactive tool combines maps with other visual presentations such as bubble charts and a parallel coordinates plotter, allowing users to select and analyse groups of regions of interest. Stories can be created that can then be shared with other users. The development of *OECD eExplorer* is the result of the collaboration between OECD and the National Centre for Visual Analytics (NCVA) in Norrköping (Sweden), led by Professor Mikael Jern.

VII. Increase in demand – new challenges

21. Table 1 below shows the increased interest in using the OECD SIS software from other organisations in 2009 and 2010. This increase in demand has led to new challenges in the technical implementation, project management, administrative and legal aspects.

Table 1.

	Organisation	Status
2006	UNESCAP	Using in production
2007	IMF	Using in production
2008	Statistics NZ	Using in production
2009	Australian Treasury Bureau	Evaluating
	Australian Bureau Statistics	Preparing for production
	US Bureau of Labour Statistics	Evaluating
	CBS Netherlands	Evaluating
	Estonia Statistics	Evaluating
	European Union	Preparing for production
	FAO	Evaluating
	iStat (Italy)	Preparing for production
	International Transport Forum	Preparing for production
	UNECA	Evaluating
	World Bank	Evaluating
2010	ONS (UK)	Requested information
	StatCan (Canada)	Requested information
	The Economist	Requested information

22. There are a number of challenges involved in finding the best way to move forward in sharing software. These include:
- Finding the right scenario for managing shared software from a centrally managed system through to a consortium approach
 - Choosing between centrally managed 'owned' software or open source
 - Finding the right balance on cost recovery
23. Despite the clear advantages of having organisations collaborate on projects to share software, such joint ventures can result in the increased complexity of project management depending on the role of the various partner organizations.
24. OECD are currently reviewing how subsequent ongoing partnerships with other organizations can be managed and are exploring the different options ranging from the most complex (consortium of development partners) to the most simple (single agency coordinating development activity and distributing code) along with all the options in between.
25. As well as the technical challenges of the IT implementations and the complexity of managing a single project across multiple organizations, the OECD is now also facing new obstacles on the legal and administrative aspects of collaborating and sharing its software.
26. Since the start of the collaboration exercise, many of the technical obstacles in sharing software have been overcome by making the applications generic and simplifying the installation procedures. In addition, project management complexity has been avoided by distributing software to organizations who can adapt it as they wish rather than operating as a consortium managing a single project.
27. The new challenges are now in the legal and administrative domains: The OECD originally aimed to recover some cost from partner organizations in order to re-invest in resources that could be used to provide support in installing, configuring and adapting the software. The increase in the number of partners has meant that this can no longer be managed within the framework of the organizations rules

as the OECD cannot charge for services. The legal department is now reviewing how this should be managed which is complex and time-consuming.

VIII. Conclusions

28. Based on our experiences sharing Statistical software with other organisations, OECD firmly believes this to be a good thing for reasons of cost savings and promotion of international standards, in particular SDMX. Despite the complexities of collaborating on projects and the administrative and legal aspects, the advantages strongly outweigh any disadvantages and we feel that this approach should continue to evolve and be adopted and promoted more widely. The OECD seeks the inputs and experiences of other national and international organisations to help determine how to best manage software sharing in the future.
-