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 UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Geneva, 9-11 February 2004)

Topic (ii): Metadata interchange

ACHIEVING EASIER INTERCHANGE OF DATA AND METADATA: THE NAWWE PROJECT

Contributed Paper

Submitted by the Australian Bureau of Statistics, Australia[1]

SUMMARY

Along with many national statistical offices, the Australian Bureau of Statistics provides data and metadata to international organisations, such as the OECD. We also have many domestic clients of ABS information. Numerous proprietary and specialised data formats have been the basis for this service. With the emergence of XML, ABS saw the opportunity to reduce the number of different formats. It was only natural to volunteer for the OECD NAWWE pilot when the opportunity arose as their proposals aligned with our strategic directions.

This paper is a case study of our experience so far. It covers the background to the project including what is NAWWE and the business benefits of the project; what ABS has done to set up the pilot and the particular problems that we had to address; and our view on the preferred approach to the use of XML for the interchange of data and metadata, taking into account our protocols about embargo of statistics and the management of releases of statistical information.

[1.] Prepared by Graeme Oakley and Jeremy Michell

I.   INTRODUCTION

1.   The Australian Bureau of Statistics (ABS) has volunteered to participate in a pilot project to develop a new way of exchanging data across statistical organisations. The fundamental idea behind the pilot is to have each organisation expose a set of National Accounts data on their web site in a standard XML format. This data can then be made available to other statistical organisations via a set of web services maintained by the OECD. The OECD have called the project - NAWWE (National Accounts World Wide Exchange).

2.   Currently the ABS fulfils its statistical obligations to the OECD by 'hand filling' spreadsheets supplied by the OECD. The data is represented as time series. Once completed these spreadsheets are sent to the OECD for compilation with data supplied by other National Statistical Organisations (NSOs).   This process has a number of disadvantages:
- costly and time consuming for statistical organisations to fulfil their statistical obligations to OECD.
- data is only accessible to OECD, not to other interested NSOs.
- the supplied data is not well described.

3.   The XML format which will be used to encapsulate the time series data has not been finalised. ABS understands that the European Central Bank is working on a standard schema for time series statistics called GESMES. It was envisaged that this schema would be used as a starting point for the OECD time series schema, and may well be adopted fully for this purpose. We understand that the GESMES XML schema is still being developed and so far it has not reached the stage where example XML files can be produced. [ABS has had access to an early version of the GESMES XML schema and has developed a transformation to produce a file from our internal schema that is very close to this early version.] More recently, OECD documentation has mentioned an XML schema called SDMX version 1.0. Hopefully, these are common and we don't end up with two competing XML schemas for time series data. Also, we have heard mention of work on a more "cube based" XML. Are we heading towards a case of "time series" vs "cube" XML?

II.   THE PROJECT

4.   The ABS project is to develop a process which allows the extraction of the required National Accounts data from the ABS Information Warehouse (ABSIW), transform it into the OECD XML format, and load it to the ABS web site.

5.   The project comprised a number of streams of work. These are:
- setting up metadata, and loading National Accounts data to the ABS Information Warehouse to support the OECD data requirements.
- develop / refine mechanism to extract required data.
- develop process to transform the extracted data into the OECD XML format using the OECD preferred approach of an EXCEL spreadsheet and then running a Macro.
- develop mechanism to place the XML files on the ABS web site, subject to release management and embargo policies.

6.   The NAWWE objectives are to:
- store National Accounts data on ABS web site and allow the sharing of data via OECD XML with other National Statistical Organisations.
- encourage NSOs to use this XML file for sourcing own data on National Accounts.

- provide web services to access the National Accounts information of participating countries. This web service will be maintained by the OECD.

7.   The ABS objectives are to:
- ensure that all relevant National Accounts data and associated metadata (including time series information) is available on the ABS Information Warehouse.
- develop mechanism to extract and transform data to OECD XML format for time series that would be re-useable for other clients.
- develop the location on the ABS web site for XML files and the 'plumbing' to place them on the site subject to embargo rules.

8.   The following diagram has been sourced from a presentation by Russell Penlington of OECD to the OECD Meeting of National Accounts Experts in October 2003. It shows the architecture and data flow envisaged by the OECD. Note that their solution is based on national statistical agencies producing EXCEL tables from which XML data files are created for transfer. This approach is very dependent on the EXCEL table format remaining static. Also, the OECD propose using SDMX based schemas.
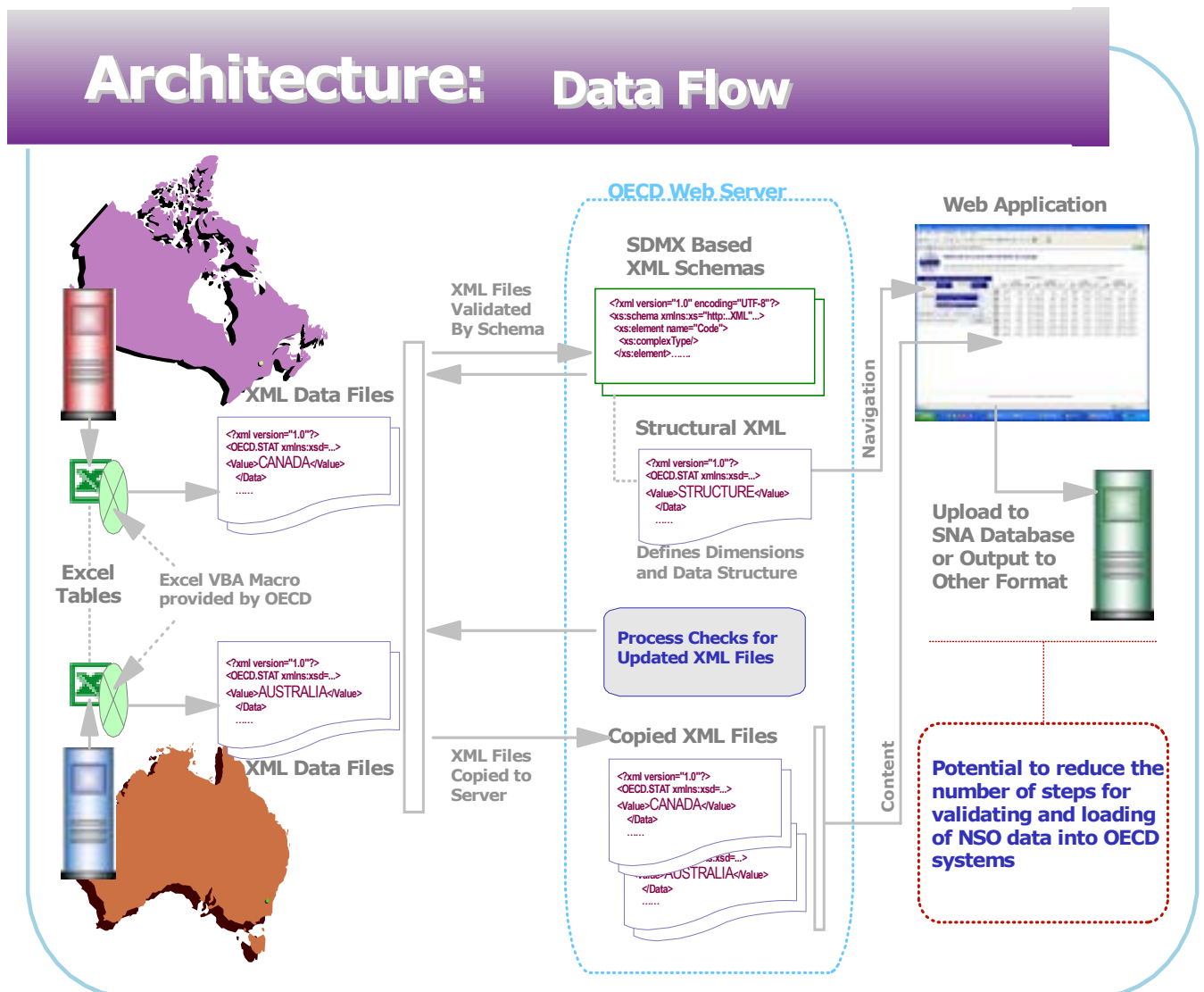
9.   The following picture is part of the spreadsheet that is populated by the ABS before creation of the XML file via Macro.

Table:	0101
Sheet:	A-COP95
Label:	Gross Value Added at Basic Prices and Gross Domestic Product at Market Prices

KEY INFORMATION		SENDER		OECD / EUROSTAT FOOTNOTES	
Reporting Country:		Name:		Footnote 1:	
Prices:	COP95	Telephone:		Footnote 2:	
Field_1 (Transaction):	row 23	Fax :		Footnote 3:	
Field_2 (Industry):	row 24	**VERSION**		Footnote 4:	
Field_3 (Not used):		Version:		Footnote 5:	
Field_4 (Not used):		Update:		**SPECIFIC ELEMENTS**	
Time:	col 2	Date :			
Unit:	NAC	**DATA START**			
Unit Multiplier:	6	Row :	26	**SENDER FOOTNOTES**	
Flag:		Column:	3	Footnote 1:	
Adjustment Type:		**CONTROL**		Footnote 2:	
Adjustment Method:		Row :	23	Footnote 3:	
Observation Type:	TOT	Column:	2	Footnote 4:	

Quarter/Year	Time Code	Gross Value Added by Industry (Combination of Letters of ISIC Rev.3 / NAC						Total A6
Code of Transaction		TRB1G	TRB1G	TRB1G	TRB1G	TRB1G	TRB1G	TRB1G
Code of Industries		AYA+AYB	AYC_AYE	AYF	AYG_AYI	AYJ+AYK	AYL_AYP	TA6
1	2	3	4	5	6	7	8	9=3+...+8
Year 1990	1990							
Year 1991	1991							
Year 1992	1992							
Year 1993	1993							
Year 1994	1994							
Year 1995	1995							
Year 1996	1996							
Year 1997	1997							
Year 1998	1998							
Year 1999	1999							
Year 2000	2000							
Year 2001	2001							
Year 2002	2002							

10.   In technical terms, the approach is based on XML and its associated technology standards. NSO's would post data on their web site using a standard XML based format. The files can be password protected or not. Then a directory of the corresponding XML files URLs for each country is made available, for example, on the OECD site. The OECD would make available a "web application" to extract data from the system using parameters, such as the country identifier and the national accounts standard SNA93 code. The OECD suggests that each NSO can also use the same XML file as a central source of data for its own dissemination for all formats and media. The NSO can easily mix data from other countries and from International organisations for international comparisons. The ABS cannot see, at this time, that this is the way that we would go about dissemination in the future. Certainly we could use the common format to obtain data from other countries for comparison purposes, but we would prepare dissemination products from the 'richer' information warehouse environment, ie not be limited by a specific subject matter area format and simple schema definition.

11.   Here is a representation of the OECD schema that is produced using the product, XML Spy.
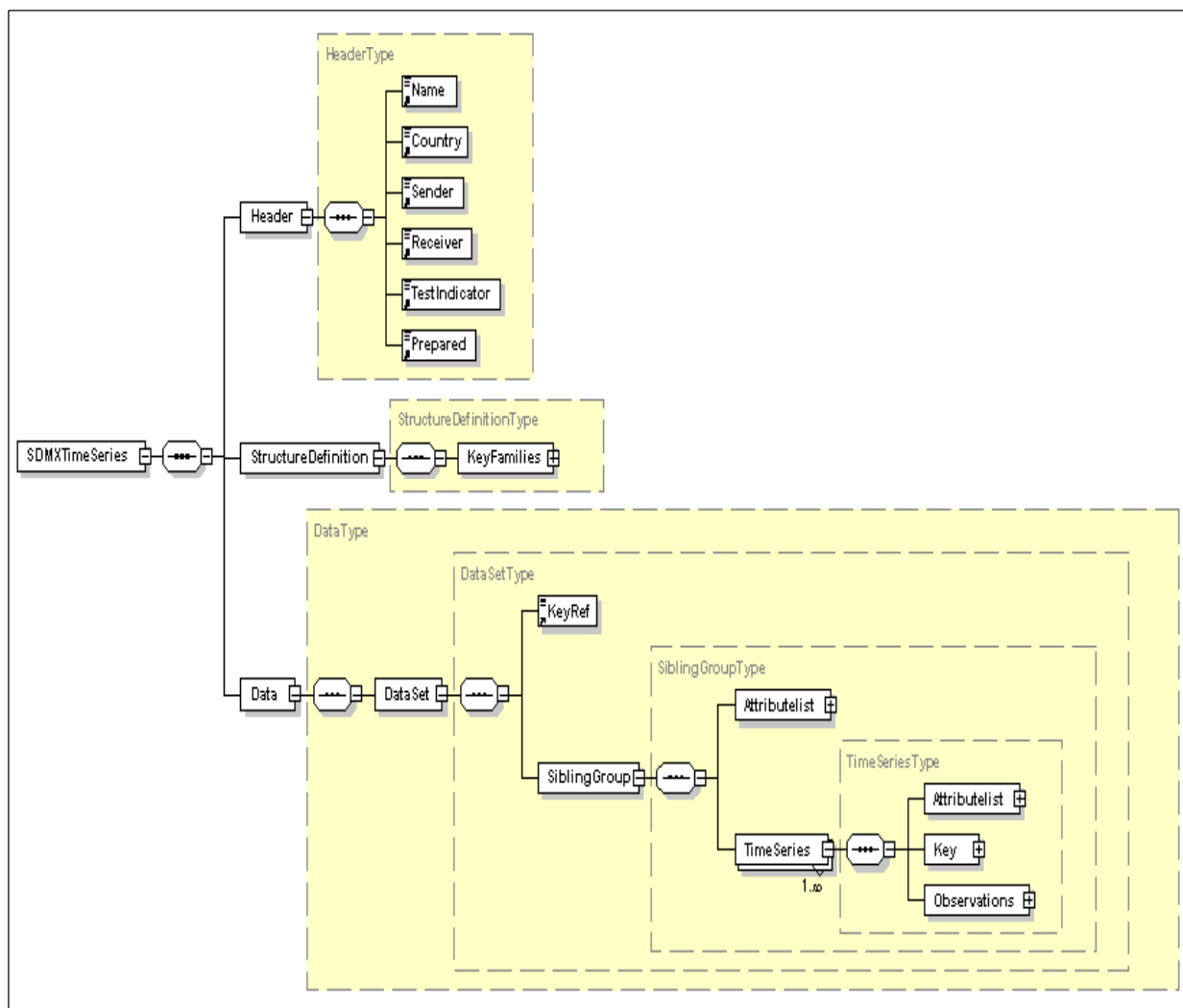
III.   PROJECT STATUS

12.   OECD has developed an EXCEL macro that will allow NSO's to produce XML files from their existing EXCEL files. At present the XML tags in the proposed XML schema are specific to NAWWE but they could be adjusted to statistical standards when they become available. The next steps from the OECD viewpoint are to:
- modify and refine the XML schema as SDMX version 1.0 eventuates
- enable notification of XML file import status (NSOs and OECD staff)
- provide XML output functionalities (EXCEL and database upload)
- integration of user functions (validation, output, upload).

13.   In November 2003, ABS put up XML files on its web site using the OECD EXCEL macro applied to data entered by hand into the relevant standard spreadsheet. Whilst most of the Australian national accounts data that is disseminated comes from our Information Warehouse, there were a number of series provided to OECD that had not been loaded, and there were a few issues with setting up the dataset to create the specific OECD spreadsheet format. National Accounts staff have worked at dealing with these issues on and off over the 3 or 4 months prior to November 2003 as time allowed during a particularly intensive processing and dissemination schedule. We took the clerical approach in the interests of being able to test out the infrastructure. The next steps from the ABS viewpoint are to:

- complete the loading of national accounts data for OECD to our Information Warehouse.
- generate XML files according to the OECD schema from our internal time series XML schema that delivers information from our warehouse. In due course, this translation would be switched from the OECD schema to the SDMX schema.

IV. ABS EXPERIENCE WITH PILOT PROJECT

A. Current Process

14. In the current process, data is hand-entered into an EXCEL spreadsheet by the National Accounts area and sent back to the OECD. The same data is loaded to the ABS Information Warehouse (ABSIW) and used to produce statistical products (see Figure 1).

Pros
- Relatively unsophisticated method which allows manual intervention

Cons
- Time consuming to enter data into EXCEL template
- Possibility of errors in the transcription from FAME to EXCEL, and potential for discrepancies with other disseminated data because the information is not sourced from the ABSIW.

Figure 1 : Current Process

OECD

FAME Time Series Database

Hand Entered

Excel File

ABS Information Warehouse

Publications
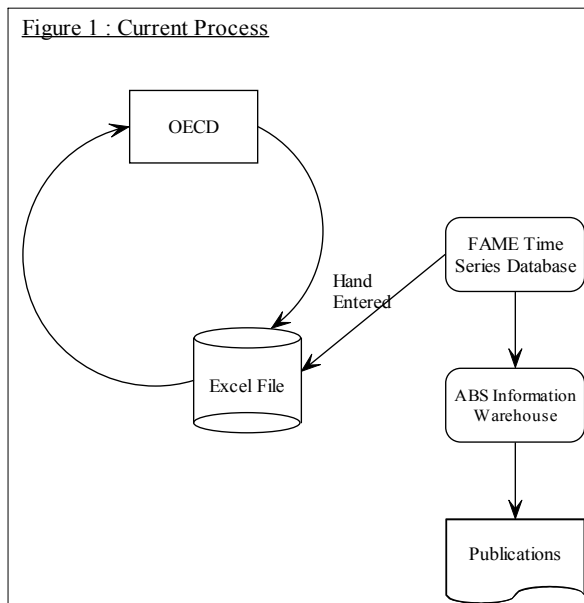
B. Intermediate Process (basis of files provided in November 2003)

15. In the intermediate process data is hand-entered into an EXCEL spreadsheet. Once this has been completed, an EXCEL macro is run against the spreadsheet to produce a relatively simple XML output file. This output file is placed on the ABS web site to allow the OECD to access it (see Figure 2). It is envisaged that the OECD will provide a services interface to allow other entities to access ABS data in conjunction with data provided by other countries.

Pros
- Data provided in a slightly richer (more descriptive) format.
- Data accessible by other organisations apart from OECD.

Cons

- Time consuming to enter data into EXCEL template
- Possibility of errors in the transcription from FAME to EXCEL, and possibility of differences with other disseminated data because not all outputs are sourced from the ABSIW
- Extra manual step to convert EXCEL to XML
- Maintenance required if spreadsheet format or XML schema changes.

Figure 2 : Intermediate Process
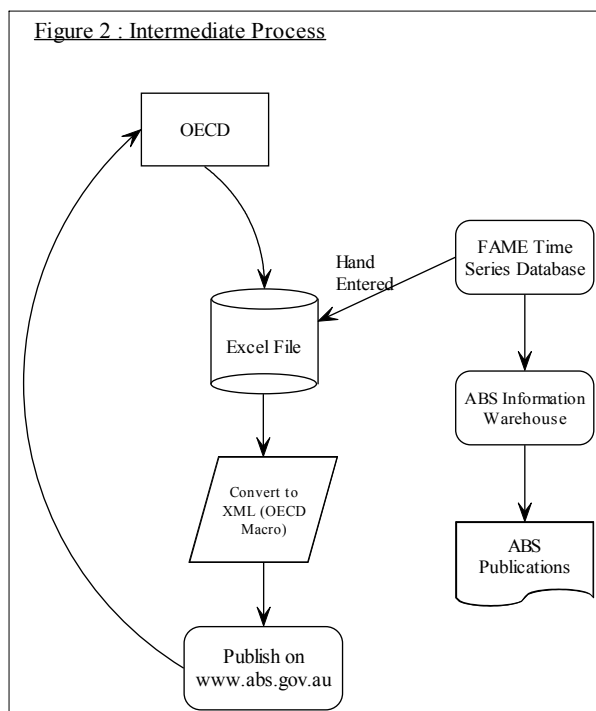
C.   Longer Term Process

16.   In this longer term process, data is loaded to the ABS IW as it was in the previous processes. An XML file is produced from the ABS IW in a proprietary ABS XML format. This XML is then transformed to an international standard XML schema (like GESMES or SDMX XML schema) and placed onto the ABS web site to be accessed by the OECD (see Figure 3). It is envisaged that the OECD will provide a services interface to allow other entities to access ABS data in conjunction with data provided by other countries.
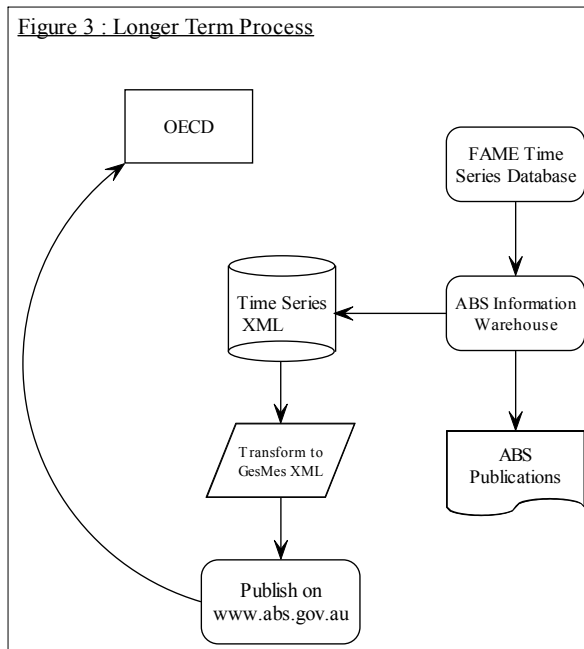
Pros
- Data provided to OECD is sourced from same place as other ABS disseminated products.
- No manual intervention required, so no scope for transcription errors.
- Data provided in a very rich (more descriptive) format.
- Data accessible by other organisations apart from OECD.

Cons
- Susceptible to changes in international XML schema.

Figure 3 : Longer Term Process

V. ISSUES FOR CONSIDERATION

17.   The ABS raises the following issues for consideration as a result of the experience so far.

   i. The intermediate step of an EXCEL spreadsheet in a specific format seems to be restrictive in terms of future changes. It seems that any change to the spreadsheet will require a change to the EXCEL macro and possibly the XML schema. ABS preference is for the information to be provided directly in an XML schema without the intermediate step of EXCEL spreadsheet format.

   ii. The spreadsheet layout is very specific to OECD requirements and national accounts. Its 'style' is different to spreadsheets placed on the ABS web site so we would be unlikely to make it available to other clients because we then abandon a 'common look and feel'. Is this an issue for other agencies? [This highlights the importance of the medium/long term move to direct XML, because then we can all share the same XML but render/present it (for non OECD purposes) into whatever spreadsheet (or other) format meets our standard 'look and feel'.]

   iii. Which XML schema will be used for time series information. It is not clear what is the relationship between SDMX work and ECB with GESMES. We need an agreed treatment for identification of series; annotations for cells, individual series; descriptions of data elements etc.

   iv. What are the plans for extension to other data sets? ABS is aware of an initiative related to international trade data and involving OECD and UNSD. ABS would prefer a common XML schema for the provision of all data requirements to international statistical organisations.