

**Unclassified**

**STD/SIMS(2004)5**



Organisation de Coopération et de Développement Economiques  
Organisation for Economic Co-operation and Development

**09-Mar-2004**

**English - Or. English**

**STATISTICS DIRECTORATE  
STATISTICAL INFORMATION MANAGEMENT AND SUPPORT DIVISION**

**STD/SIMS(2004)5  
Unclassified**

**Experiences, Aspirations and Strategies of the Australian Bureau of Statistics related to Data and Metadata Exchange**

**OECD EXPERT GROUP ON STATISTICAL DATA AND METADATA EXCHANGE**

**1-2 April 2004  
Château de la Muette, Paris**

**JT00159548**

Document complet disponible sur OLIS dans son format d'origine  
Complete document available on OLIS in its original format

**English - Or. English**

## **EXPERIENCES, ASPIRATIONS AND STRATEGIES OF THE AUSTRALIAN BUREAU OF STATISTICS RELATED TO DATA AND METADATA EXCHANGE**

Data Management Branch  
Australian Bureau of Statistics

### **A. Introduction**

1. The Australian Bureau of Statistics (ABS) is very interested in both:
  - sharing of data and metadata with international organisations, and
  - sharing of data and metadata within a National Statistical Service (NSS).
2. The NSS initiative aims to improve the quality and breadth of information available to Australians. While the ABS collects and manages statistics both nationally and on behalf of the constituent states and territories, it is not the source of all statistical data (including statistics which can be generated from administrative datasets) which are of value to Australians.
3. While more characteristic of a provider/customer relationship than sharing among partners, the ABS is also, of course, very interested in improved and more consistent means of delivering data, associated with appropriate metadata, to the end user community. It is expected that aspects of the protocols and standards developed to promote sharing among international and national partners will, directly or through extension, also benefit some aspects of supporting the end user community. At least one key organisation in the user community has already expressed interest in the possibility, for example, of receiving data in the SDMX XML format in future. More generally, sharing of standard transforms based around these standard formats may in future provide a low cost means of presenting data on a consistent basis in a final format that best meets a particular user's needs.
4. This paper looks at the current situation in regard to sharing data and metadata with international organisations. As the NSS is still in the process of being established, "current situation" is not explored in detail for it.
5. The paper then looks at the overlapping ABS aspirations and strategies in regard to the sharing of data and metadata both with international organisations and within the NSS.
6. A summary is then provided of ABS experiences with a number of case studies in regard to sharing of data and metadata with international organisations.

### **B. Current situation**

#### *Sharing of data and metadata with international organisations*

7. As per the following table, the ABS provides selected data on a regular basis to at least ten different international agencies. Some data is supplied on a monthly basis, other reports are on a quarterly or annual basis. More than a hundred "regular" data reports of various kinds are dispatched each year. This is in addition to nearly fifty "one off" data reports provided to international organisations and National Statistical Offices (NSOs) last year.

Organisation	Number of distinct questionnaires	Total number of reports per year
Economic Commission for Europe	1	4
Economic and Social Council for Asia and the Pacific	3	17
Food and Agriculture Organisation	7	7
International Energy Agency	1	4
International Labour Organisation	3	3
International Monetary Fund	2	16
Organisation for Economic Co-operation and Development	4	7
Secretariat of the Pacific Community	1	4
United Nations (Not further defined)	6	23
World Tourism Organisation	6	23
<b>TOTAL</b>	<b>34</b>	<b>108</b>

8. Each data report currently tends to have its own unique formatting requirements. This often applies even when reporting different fields of statistics to the same organisation.

9. The specific reporting formats often require that, once suitable data has been extracted from a database using a standard delivery format, the data is extensively “reshaped” through spreadsheet manipulations and/or “cut and paste” operations. As well as being labour intensive, the process is error prone. Most reports are produced quarterly or annually so providers seldom bother to automate the manipulation process in a controlled manner (eg via macros). The knowledge of what was done last time is often lost, especially when staff move on. At a minimum this means a lot of time needs to be spent validating the report before it is dispatched. At worst, it means the report may contain data errors or differ without due explanation from estimates that the ABS has published elsewhere.

10. There are also concerns in some cases that the metadata being provided along with the data is insufficient due to limitations in the reporting format and/or the methods of producing the customised outputs. Sometimes data is supplied on a regular basis using specifications that were submitted on a one off basis a long time ago. There is no metadata that the ABS or the receiving agency can check each time to ensure that the data remains aligned with the most recent version of those specifications.

11. Finally, having extracted, collated and transformed data to produce the report, there is little - if any - additional value to be gained from this work. The same output cannot be used to meet other reporting needs. Even where the data is generally releasable, the “patchwork quilt” of specialised reporting formats means it is not attractive to provide the outputs to a wider audience via the ABS web site where there is the aim of maximising consistency in “look and fee” and of ensuring suitable metadata is associated with every output.

12. Similar issues can arise when providing data in a specialised format on a regular basis to a couple of other government agencies within Australia.

### **C. Aspirations and strategies**

#### *Sharing of data and metadata with international organisations*

13. The ABS recognises at least two types of standardisation which can assist in sharing data and metadata. The first is **standardising the format for the interchange**, such as using a common XML schema. The second is **standardising the content of the interchange**, such as ensuring that each agency has populated data for each data element based on a standard definition of that data element. In the case of interchange with international organisations, there has traditionally been a focus on the second type of standardisation, with each agency reporting data against a well defined framework (eg System of National Accounts). Interchange formats themselves, in the past, have been less standardised across the international community.

14. The ABS believes that, with the first type of standardisation, the common interchange format should be metadata rich - even if not every attribute is mandatory for every purpose. One reason for this is that well described metadata can assist in determining whether the second type of standardisation (content) has actually been achieved or whether subtle definitional differences persist.

15. In terms of the first type of standardisation, the ABS has developed a detailed XML schema for statistical data (as data cubes and/or time series) and metadata. This schema is intended for use within the ABS, as a standard interchange format between repositories and applications.

16. Rather than directly supporting extraction of data and metadata in a range of different spreadsheet and time series output formats, the ABS would like to be able to extract data and metadata in this 'internal' XML format against which transforms can be run to create final outputs in a limited range of formats. One favoured format, or set of formats, will be XML schemas defined by international organisations. To date these schemas have tended to be less extensive and more purpose specific than the ABS schema. This should work well as long as the ABS schema, at a minimum, supports all the concepts found within the other schema.

17. Some international initiatives have pointed to benefits in terms of “respondent burden” to NSOs of only having to extract data once and then have it shared in a common format between multiple international organisations and other interested parties. Benefits in this area are recognised, but from an ABS perspective it would not be too onerous to continue to have to extract different sets of data of interest to different organisations as long as these outputs could be produced in a common format. XML is seen as providing great opportunities for achieving this as:

- not every element of the schema needs to be of relevance to every extract (the organisation receiving the XML can ignore what they don't need);
- the schema is extensible (at some cost to supplying organisations to populate new elements) if additional reporting needs arise over time.

18. The ability to generate output in this format in an automated manner not only saves time compared to “hand crafted” reporting formats, it also significantly improves quality assurance.

19. The ABS is very keen to participate in development processes, including case studies, in this area. Benefits to the ABS include the ability to:

- learn from the perspectives and experiences of international organisations and other NSOs in regard to data and metadata exchange;
- influence the design of the schema to be able to support as many reporting purposes as possible without adding undue complexity;
- check whether the proposed international schema contains all the concepts that appear relevant;
- influence the design of the schema to ensure it incorporates appropriate levels of metadata to define and describe the data;
- see whether there are concepts proposed for the international schema that should be incorporated in the ABS schema.

20. The ABS recognises that there are many steps between arriving at an agreed format for one particular reporting requirement and having that format adopted much more widely for international reporting purposes. One challenge is ensuring that, without unduly compromising its initial purpose, an interchange format developed to accommodate a particular set of content is equally able to accommodate content from other fields of statistics. Nevertheless, with current levels of communication and cooperation between organisations at this level, achieving such an outcome from a sound starting point should be possible.

21. It will also be important that establishment of any common reporting format does not distract attention from ongoing efforts to advance standardisation of content within various fields of statistics. The latter is likely to remain a demanding and time consuming process, even if/when the former has been achieved.

22. By arriving at a format that can be used for many international reporting purposes, it is more likely that NSOs and other organisations will become interested in the format for other purposes. XML, once again, offers a range of opportunities in this area:

- standard transforms can be defined to translate the data into spreadsheet, and other, formats to make it easy for users to access and understand data in the standard format even if they can't consume the XML natively;
- users who require data in a specialised format for internal processes can write their own transforms to be used against many datasets in the standard format (assuming the standard format is able to address the mandatory data concepts in the specialised format).

23. Under these circumstances it would make a lot of sense for the ABS to make available on the web site, in the same format, the releasable data it had provided to international organisations.

#### ***Sharing of data and metadata within a National Statistical Service (NSS)***

24. A common interchange format of the type described above could prove valuable in promoting exchange across the NSS in the medium term.

25. It is important to recognise, however, that there are likely to be some differences between some partners in the NSS and the partnerships formed between NSOs and international organisations.

**Characteristics of international reporting by NSOs**

Managing and exchanging statistical data and metadata is "core business".

Data and metadata sought by international organisations is typically based on well defined international frameworks of concepts and classifications.

International organisation determines what data they require.

Each NSO provides the data specified in the reporting requirements.

International reporting typically seeks information on a limited set of "core measures"

**Characteristics of NSS partners**

Managing and exchanging statistical data and metadata may be secondary to a wider mission.

Some data to be presented by, and exchanged between, NSS partners may not be based on standardised concepts, or may be based on conceptual frameworks other than those employed by the ABS and other NSOs.

NSS partners determine which data they wish to make available.

Each NSS partner makes available a unique set of data.

The diversity of subject matter, and the level of detail, of data to be presented and exchanged is likely to be much greater than that typically associated with international reporting.

26. The characteristics of the NSS make it particularly important that presentation and interchange formats provide sufficient metadata to be able to determine the definitions, and areas of commonality and difference, of data reported from heterogeneous sources. It is also important, however, that the requirements are not so complex and demanding that they act as a disincentive for partner agencies to voluntarily contribute content to the NSS.

27. Although there are various initiatives to achieve standardisation of content in specific subject matter areas (eg Health, Education), achieving a standardisation of content across the NSS as a whole would appear to be a distant, or impossible, prospect. Encouraging partners within the NSS to express their data in a common, metadata rich, format may, however, promote extension of, and mapping between, common definitional frameworks. Data from different sources will be able to be compared and confronted more effectively, allowing for the identification of commonalities and differences, with the option of seeking to harmonise the latter.

28. One of the first steps is for NSS partners to provide some basic metadata in a standard format to describe their data sources. This advertises the availability of the data and provides contact details but does not attempt to force the data itself into a standard presentation and exchange format at this early stage.

29. Best practice advice in managing data and metadata is also being shared. For example, the Australian Institute of Health and Welfare's (AIHW) work in providing a "Knowledgebase" of Data Elements and Data Element Concepts is promoted. While recognising a "one size fits all" mentality is not appropriate, this work encourages some level of commonality in thinking which should make it easier to move toward commonality of exchange formats in the future.

30. Finally the ABS is supporting initiatives related to the establishment of National Data Network for research purposes. This is seen as another opportunity to progress work on standardising the presentation of data and metadata and to underline the value of standardisation.

31. In the medium term, as a viable interchange format emerges, it is expected that some partners who don't have well developed processes for managing and exchanging metadata internally to their organisation may adopt the standard format for that purpose. Other partners may map to the standard format only for external presentation and exchange purposes.

**D. Experience to date**

32. The ABS has had a number of experiences working with generalised schemes to share data and metadata via XML. The ABS has worked with those responsible for the development of the Statistical Data and Metadata Exchange (SDMX) format to ensure that the concepts described in the ABS XML Schema support those required in

SDMX. The ABS is also participating in the National Accounts World Wide Exchange (NAWWE) project with Statistics Canada and the OECD.

33. As mentioned above, the ABS developed a detailed XML schema for statistical data and associated metadata to facilitate the sharing of information between the central store of publishable data (the ABS Information Warehouse) and tools to compile and display this publishable data. The schema was set up to support both sets of individual time series where the data has two dimensions (time series and a time component) and data cubes where the data can be described by any number of dimensions. The initial implementation of the delivery of this schema from the ABS Information Warehouse was via the time series format. The resulting XML was then used to produce a spreadsheet product for the dissemination of time series data.

34. At this time we became aware of the work going on in the European Central Bank and the OECD to replace the existing GESMES format for the sharing of time series data with an XML implementation (SDMX). Some discussions occurred around how the two formats (SDMX and the ABS XML Schema) described the various concepts used to describe time series data. The development of a single schema for the transmission of data and metadata between international statistical agencies was seen by the ABS as a real opportunity to improve the quality (in terms of comprehensive information provided with the data delivered and reducing the need to customise the data delivered to various agencies) of data, as well as reducing the burden on the organisation of providing this data in various formats. Once some solid examples of the SDMX format were available, the ABS went about developing an XML transform to convert from the ABS XML schema to the SDMX XML schema. As part of this work some deficiencies in the ABS schema were identified. The main one was that the time series was described textually without any solid links back to the underlying classifications, which meant that CodeLists components of the SDMX format was not available. Subsequent work was done to fill out the ABS XML schema to include this more comprehensive information. The ABS is now waiting on SDMX Version 1.0 before finalising the XML transform.

35. The ABS volunteered to take part in the OECD NAWWE project which aims to have contributing agencies provide national accounts data on their own Website in SDMX format. The OECD will then provide a service via its Website to allow other interested parties access the data / metadata of all participating agencies from a central point. Currently the project aims to create SDMX XML files by running a process over the spreadsheets that are currently supplied to the OECD. In the future the ABS aims to produce the National Accounts information in SDMX format via a transform on the ABS XML data delivered directly from the ABS Information Warehouse. This will reduce the manual labour and associate risk of transcription errors that are associated with hand entering data into the OECD spreadsheets.

## **E. Conclusions**

36. Early experiences have realised some of the benefits that the ABS was seeking through participation in case studies. This has included helping us test and improve our internal XML schema and align it with wider thinking.

37. The work has maintained a wider focus, ensuring the format developed through each case study is not limited to interchange of just the specific set of statistics which are the initial subject for the study.

38. The ABS has received excellent value from its participation in case studies to date. We believe that while relatively slow and incremental, the case study approach, particularly where sponsored by high profile international organisations, offers some of the best prospects of developing an effective and practical interchange format for statistical data and metadata which will find high levels of acceptance on a global scale. Once defined and accepted, such an interchange format is also likely to prove valuable for more local and for bilateral purposes, especially if supported by a range of standard transforms.

## **F. Discussion Points**

39. To what extent do other agencies believe that:

- a standard format developed for interchanging a standardised set of statistics can be applied to interchanging data for other fields of statistics;

- a standard interchange format designed for homogeneous international reporting can assist in exchange of heterogeneous data and metadata between partners within a National Statistical Service, and with delivery of data and metadata to end users.

What are the deployment and operational issues/problems that might need to be solved?