

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

Workshop on Data Dissemination Systems
(Geneva, 12 May 2008)

SUMMARY OF THE WORKSHOP

1. The UNECE Workshop on Developing Data Dissemination Systems was held in Geneva on 12 May 2008. Participants from the following countries attended the workshop: Albania, Australia, Azerbaijan, Belgium, Bosnia And Herzegovina, Brazil, Czech Republic, Estonia, Finland, Georgia, Greece, Israel, Kyrgyzstan, Latvia, Luxembourg, Moldova, Montenegro, Poland, Russian Federation, Slovakia, Slovenia, Sweden, Switzerland, Tajikistan, Ukraine, United Arab Emirates, United Kingdom, United States of America, and Uzbekistan. The European Commission and the IMF were represented at the workshop.
2. Mr. Steven Vale, Database Coordinator, UNECE, Statistical Division, welcomed the participants and explained the aims and objectives of the workshop. The growing importance of the Internet in dissemination of statistics was in the focus of the meeting. The workshop was organized to provide an opportunity for participants to share experiences in web-based dissemination. Mr. Vale also chaired the meeting.
3. The agenda of the workshop consisted of the following substantive topics:
 - (i) Current data management and dissemination tools and systems;
 - (ii) Systems and tools - UNECE experiences with PC-Axis;
 - (iii) Practical applications of dissemination systems in national statistical offices;
 - (iv) The scope for future joint development of data management and dissemination tools.

FURTHER INFORMATION

4. Presentations and all background documents for the meeting are available on the website of the UNECE Statistical Division (<http://www.unece.org/stats/documents/2008.05.dissemination3.htm>).

I. CURRENT DATA MANAGEMENT AND DISSEMINATION TOOLS AND SYSTEMS

5. The participants followed presentations on principles of dissemination; good and bad examples of current practice and introduction to the concept of data cubes.
6. The attention of the audience was drawn to the need to move from passive dissemination, making data available in the hope that someone would use them, towards active marketing, building relationships and convincing potential users to use official statistics. Statistics have to be used in order to be useful, so statistical offices have to consider user needs, and explain the value of information that already exists. Participants noted the importance of statistical databases for dissemination, as well as for the efficient production of statistics.
7. The workshop participants also considered the good and bad current practices of the Internet based dissemination. The survey of websites of national statistical offices showed that 51.8% of countries used static pages (HTML, Word, etc.), 21.4% used spreadsheets (mostly Excel) and remaining countries used various types of dissemination databases. The following points we made

- Static HTML pages can be loaded rather quickly, and they do not need any special software, plug-in on either server or client side. However, they do not offer many possibilities for reuse of data and customising the output to the user needs. These pages also quite labour intensive for updating, in principle they should be updated manually.
- Excel spreadsheets represent the most commonly used format among data users. Users can download data and customise them. However, Excel software can only be purchased with the MS Office suite (which is not cheap). Users also have to download megabytes of Excel files to get a couple of figures.
- National databases provide solutions tailored to national needs. Users can download only data of their interest and manipulate them on the client side, but often also on the server side. However, there are considerable costs to such developments, and there is evident duplication of efforts among countries.
- PC-Axis is the most popular data dissemination software. It allows data extraction and customization by users. There is a strong user community behind the development and further extensions of PC-Axis. Many popular add-ons are available as free downloads. However, there is a (small) license fee for statistical offices using PC-Axis. While PC-Axis is shared by several statistical offices, the burden on translation of interfaces to the national language(s) is on the country that decides to use this software.
- Statbank/PC-Axis, SuperWEB, DevInfo and OECD.Stat were mentioned as other data dissemination systems currently in use.
- SDMX format became the key format for exchange of statistical data with a view of their re-use in the receiving user's system. Therefore, it is important to implement export of data in this format.

8. Multidimensional data cubes represent a flexible means of storing and presenting of statistical data according to unique combinations of several classifications. The following points were made in the discussion:

- The current software usually allows for as many dimensions as needed. However a smaller number of dimensions make data selection and understanding easier for users – preferably three dimensions at a time. An example quoted was: country, education attainment, age group and year;
- Data cube design requires some experience. The best verification is through practice, and a few trial and error cycles may be needed. This was documented by experiences of several statistical offices present at the workshop.

II. SYSTEMS AND TOOLS – UNECE EXPERIENCES WITH PC-AXIS

9. The PC-Axis suite was presented at the workshop, followed by a presentation of UNECE experience in linking PC-AXIS and heritage back-end databases.

10. The functionalities and components of PC-Axis facilitate internet dissemination, as well as preparation of off-line electronic files, printed publications and graphic outputs (graphs, maps). The input to the PC-Axis can be PX files (PC-Axis own format), as well as other user databases. In future, SDMX input will be also possible. The following points were raised in the discussion:

- While PC-Axis does not have an ambition of becoming a geographical information system, the suite comprises PX-Map for cartographic representation of data on the internet in the form of choropleth maps and cartograms.
- The new version of PX-Map uses dynamic maps (using SVG format) rather than JPEG files as before.

11. The UNECE has adopted the PC-Axis solution to provide a user friendly dissemination system rich on solutions and flexible. At the same time, the UNECE focused on developing the missing pieces, in particular for linking dynamically the production and dissemination databases, defining data cubes and other tasks related to maintenance and management of the content. The UNECE developed the following tools: (i) SQL code generators for data mapping, (ii) Visual Data Mapping Designer; (ii) Visual HTML Designer and (iii) TSSL – Time Series Computation Language.

- Participants asked about possibilities for aggregation and regression over time. The UNECE addressed only the most urgent needs at present. The analytical tools are planned for a later stage, in connection with a technological re-engineering of the UNECE databases.
- The UNECE looks to the open source community before engaging into any in-house development. This avoids re-developing of what is already available.

III. PRACTICAL APPLICATIONS OF DISSEMINATION SYSTEMS IN NATIONAL STATISTICAL OFFICES

12. Statistical offices of Estonia and Slovenia shared their experiences in using PX-Web databases.

13. Statistics Estonia provided an overview of their data dissemination system. They choose the PC-Axis system for their statistical database and regional development database. The selection was based on pricing, existence of the reference group, use of PC-Axis in neighbouring countries and functionalities with respect to multidimensional cubes. Statistics Estonia also considered SuperSTAR, Beyond 20/20 and Statline. The following lessons can be learned from Estonian experience with PC-Axis:

- There is a need for training courses and for instructions how to use databases.
- It is important to put efforts in promoting the databases.
- PX-Web provides several data export formats. In the Estonian experience Excel was the most popular format for downloading of data.
- Harmonization across various subject-matter areas was a key task in implementing a unified data dissemination system. Support from senior management was needed to address cultural issues and moving from stovepipes to a corporate system.
- The examples of PX-Web pages used various look and feel at various stages of browsing and downloading of data. However, it was pointed out, that most of the PX-Web implementations have a uniform look and feel identical with the rest of the statistical office's website. The look and feel can be fully customized.
- Since the implementation of on-line databases based on PC-Axis, Statistics Estonia does not publish printed versions of data tables. The print publications issued by Statistics Estonia focus on analysis rather than on a simple data release. In general the participants concluded that the time of printed data publications is over, and that the printed publication of the future will be significantly different to once that we were used to until now.

14. The Statistical office of the Republic of Slovenia focuses on improving the reputation of official statistics through achieving satisfaction of users, improving statistical literacy and promoting democratic values with respect to access to data. All data are released simultaneously to all users through dissemination databases, and no early or privileged access is permitted. The statistical office attaches great importance to building relationships and communication. Data dissemination systems are part of a broader strategy.

- Numerous printed and on-line publications are released for users with various level of statistical knowledge. Therefore, explanations and guidance are included in these publications
- Multiple production systems and databases are used for preparation of data. The integration of the data production and dissemination systems in Slovenia is not achieved yet.
- Dissemination databases are one of the dissemination products. They are an important element in the communication pyramid. They are designed for the most demanding users of statistical data.
- Press releases issued by the statistical office are a very good link between databases, publication and the graphic material. In setting up their system for issuing press releases the statistical office learned from the existing good practices around the world. In majority of cases the access to data is provided simultaneously to the first press release. The press releases are also a good way to signal to users that data comprised in the databases are currently fresh.
- Slovenian statistics considers three major groups of users: (i) Tourists (interested in key indicators in first releases, pocket books, yearbooks, etc.); (ii) Harvesters (interested in first releases, analytical publications and databases) and (iii) Miners (interested in databases and first releases). Media would be classified as "tourists".

IV. THE SCOPE FOR FUTURE DEVELOPMENT OF DATA MANAGEMENT AND DISSEMINATION TOOLS

15. The workshop participants noted outcomes of related events, and in particular the UNECE/Eurostat/OECD Meeting on Management of Statistical Information Systems (MSIS, April 2008). The IT managers and experts participating in the meeting created a Task Force on Sharing Statistical Software. The general trend towards a collaborative development applies also to data dissemination systems. PC-Axis is one of several examples. At the same time there is a rapid development of systems and tools for data dissemination.

16. Sharing of software and components depends on the legal framework. While in some countries the open source solutions is possible, in other the sharing of technology meets legal and licensing obstacles. A suggested solution was to develop a model public license, following example of the European Union Public License (EURL, <http://ec.europa.eu/idabc/en/document/7330/5980>). The following issues have to be addressed:

- Ownership – IMF referred to their experience with taking over the OECD.Stat. From the technical viewpoint, the ownership represents a need to maintain a commonly agreed source code respecting the original donor and recipients' needs. PC-Axis consortium was suggested as a model practice in this respect.
- Language versions – the interface, and in some cases the documentation, have to be translated into the language of the recipient country.

17. Another related event was the UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS, April 2008). Lessons learned from the work session include a demand by users for more interactive data dissemination. Such an approach requires an end-to-end business process model. To this end the METIS group, led by Statistics New Zealand and UNECE, develops the Generic Statistical Business Process Model.

18. Next steps were considered. Possibilities included joining an existing network (PC-Axis), or creating a new network, as the software and components would refer also to other applications as PC-Axis. UNECE would be happy to facilitate such initiative in a substantive and technical aspects:

- Guidelines for Data Dissemination Systems;
- Mechanism for exchange of ideas, experiences and software – UNECE wiki;
- Networks for sharing development;
- Generic public license following the example of EURL and/or GPL;
- Workshops / task forces on specific issues;
- Other possible activities.

19. The following questions were raised in respect to the collaboration related to PC-Axis:

- Representatives of member countries asked what assistance the UNECE can provide in implementing PC-Axis. In response the UNECE representative stated that the UNECE is developing add-on tools for PC-Axis in a way that they can be re-used by member countries and is ready to provide them to any country interested. The UNECE does not have resources for a consultancy, but it would address this through the PC-Axis consortium.
- Member countries were also interested in costs of the PC-Axis software. There are basic fees related to the support of the consortium led by Statistics Sweden that are set with respect to the size of the country and economic possibilities. It was emphasised that PC-Axis is not distributed for profit.
- Countries interested in implementing PX-Web may wish to download (for free) the PC-Axis desktop software from the internet and try it. PX-Web is very similar, it has the same functionalities.
- The UNECE offered to facilitate a network of Russian speaking countries that wish to implement PC-Axis and/or PX-Web and share its experiences with (and results of) implementing the Russian version of its statistical database.

20. Those who would like to participate in any sort of network for sharing data dissemination software, components or experiences should express their interest to the UNECE Statistical Division.

* * * * *