**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 2008)
(Luxembourg, 7-9 April 2008)

REPORT OF THE APRIL 2008 MEETING ON THE
MANAGEMENT OF STATISTICAL INFORMATION SYSTEMS

1.   The Joint UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS) was held in Luxembourg, from 7 to 9 April 2008. Participants from the following countries attended the meeting: Bosnia and Herzegovina, Bulgaria, Canada, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Israel, Italy, Latvia, Lithuania, Mexico, Moldova, Netherlands, New Zealand, Norway, Poland, Portugal, Republic of Korea, Romania, Russian Federation, Slovakia, Slovenia, Spain, Sweden, Turkey, and United Kingdom. The European Commission was represented by Eurostat. Representatives from the following international organizations also attended: United Nations Industrial Development Organization (UNIDO), Organization for Economic Cooperation and Development (OECD), International Monetary Fund (IMF), Bank for International Settlements (BIS), and Eastern Caribbean Central Bank (ECCB).

2.   Mr. Pedro Díaz Muñoz, Director, Eurostat addressed the meeting at the opening. He stressed the importance of migration from stove-pipes to a corporate management of statistical information systems involving not only IT managers, but more importantly the senior management of statistical offices. It is important in this context that IT expertise is kept up-to-date. Mr. Diaz Muñoz also referred to ongoing Eurostat activities, in particular the statistical life-cycle project CVD ("Cycle de vie des données"), improving the statistical IT governance and standardization activities.

3.   The agenda of the meeting (WP.1) consisted of the following substantive topics:
   (i)   Governance and management strategies;
   (ii)   Statistical information systems architecture;
   (iii)   Exchange/sharing/re-use of components, common models among statistical offices;
   (iv)   User perspective.

4.   The sponsoring international organizations (UNECE, OECD and Eurostat) thanked the members of the Steering Group on Management of Statistical Information Systems for their work in preparation of this meeting.

5.   Mr. Rune Gloersen (Norway) acted as Chairman of the meeting. The preparation of the substantive work was organized by: Ms. Cathy Wright (IMF), Jolanta Stefanska (IMF) and Adam Wronski (Eurostat) for Topic (i); Ms. Karen Doherty (Canada) for Topic (ii); Mr. Marton Vucsan (Netherlands) for Topic (iii); and Mr. Lars Thygesen (OECD) for Topic (iv).

6.   The participants adopted the present report before the Meeting adjourned.

RECOMMENDED FUTURE WORK

7.   The participants appreciated the opportunity for an international exchange of experience in the field of management of statistical information systems, and recommended organizing another meeting in the future with the aim of identifying leading practices in national statistical offices.

8.   To this end the participants considered that the following themes are currently of crucial importance to IT managers in statistical offices:
*   Governance;
*   Architecture;
*   Sharing of software and components.
It was also recommended that the following topics be discussed within the three above-mentioned themes:
*   IT Infrastructure Library (ITIL);
*   Mitigating legacy systems;
*   Generic Statistical Business Process Model and its relationship to the Enterprise Architecture;
*   Keeping a user perspective.

9.   The participants agreed to launch a Task Force under the auspices of the MSIS Steering Group to prepare a report on possible future work on sharing software and components. The Task Force should deliver the first report in September 2008. The report should serve as an input for a proposed Eurostat project on sharing software and components. The following countries and international organizations volunteered to actively participate in the Task Force:
*   Canada;
*   Italy;
*   Netherlands;
*   Norway;
*   United Kingdom;
*   Eurostat;
*   OECD;
*   UNECE;
*   UNIDO.

FURTHER INFORMATION

10.   Presentations and all background documents for the meeting are available on the website of the UNECE Statistical Division (http://www.unece.org/stats/documents/2008.04.msis.htm).

ANNEX

**SUMMARY OF THE MAIN CONCLUSIONS REACHED DURING THE MEETING ON THE
MANAGEMENT OF STATISTICAL INFORMATION SYSTEMS (MSIS) 2008**

Topic (i): Governance and management strategies
Session Organizer: Cathy Wright (IMF); Discussants: Jolanta Stefanska (IMF) and Adam Wronski (Eurostat)
Documentation: Invited papers by Ireland, New Zealand, Norway and UNDP Moldova; supporting papers by Norway.

1.   The presentations and the discussion on this topic focused on the exchange of experiences in governance and management, and particularly in moving from the stove-pipe approach towards corporate statistical information systems. Issues related to e-government were also covered in the discussion.

2.   The participants discussed issues related to developing corporate systems as a replacement for the stove-pipe approach, and made the following points:
*   Reasonably small project teams cannot involve subject matter experts from all statistical areas, but are more efficient. A business intelligence system can help to gather the necessary information.
*   Convincing individual subject-matter areas to accept a corporate system may involve an individual approach to selected areas. However, a global approach would be more efficient.
*   Expected benefits include improved quality, central metadata management, as well as some financial benefits. Financial benefits may be long-term, as in the short-term they are outweighed by the migration costs.
*   Training needs should be addressed in conjunction with subject-matter areas, and the training should be business-driven.

3.   Process-based IT organization was discussed at the meeting. This comprises the introduction of functional roles and the specialisation of teams, as well as centralization and consolidation of IT. The following points were brought up in the discussion:
*   The following methodologies and tools were presented as examples:
    o   Project Management Framework (PMF);
    o   Software Development Lifecycle (SDLC);
    o   IT Infrastructure Library (ITIL).
*   According to experiences, the implementation of a process-based IT organization has brought the following challenges:
    o   Adoption of new processes is often perceived as a new administrative burden;
    o   New policy and adoption of templates is not enough;
    o   Governance & engagement becomes more complex;
    o   The process from vision to delivery can be time consuming compared to previous practice;
    o   Increased focus on building things in the right way rather than building the right things;
    o   The challenge of finding and retaining skilled staff has increased with the introduction of standard functional roles.
*   Human resources issues have to be considered when consolidating the IT organization and implementing the functional approach:
    o   New roles and specializations require recruitment and/or hiring. The example presented used mostly retraining with a minimum of new recruitment. However, changes meet some resistance, and it is necessary to build a perception of a corporate IT service.
    o   The process based organization brings benefits, if it is applied to all projects. In practice exceptions were made for very small projects. This requires discipline on the part of project managers/coordinators. A common output delivery model was suggested as one of the ways to overcome this problem.
    o   Experience suggests that when a centralised testing team is in place for all IT related activities, there is always sufficient workload. Centralised testing is part of corporate quality assurance.

      o Business analysts play an important role. Statistical offices face a challenge, as these are among the most expensive experts on the IT labour market, more costly than developers.

4.   In connection with appearance of on-line services for electronic business as well as for public services, the public expects statistical offices to join the trend. The role of statistical offices in overall e-Government activities (e-statistics) was discussed at the meeting, and the following issues were raised:

- In addition to the general motivation described above, there are particular benefits to statistics in introducing e-services:
  - o Decreased response burden on reporting units (businesses and citizens);
  - o New sources of statistical data available through on-line links;
  - o New and improved sharing of data among producers of official statistics;
  - o Improved dissemination, for example through implementing a common portal to official statistics.
- Private/Public Key Infrastructure (PKI) is often used to ensure protection for data transmitted via the Internet. Some countries indicated a lack of experience and consequently problems in developing strategies for PKI implementation. Others stressed that in countries where the use of PKI is mandatory, it results in high costs, particularly for small projects, surveys with few respondents, and specifically when private citizens are concerned.
- The introduction of e-services requires the acquisition of the necessary expertise, through training, recruitment, consultancies, outsourcing or other means.

5.   A panel discussion on this topic was organized with Adam Wronski (Eurostat) and Jolanta Stefanska (IMF) acting as Moderators. The participants in the panel were:

    - Paul Woobey (Office for National Statistics, United Kingdom)
    - Matjaz Jug (Statistics New Zealand)
    - Pedro Díaz Muñoz (Eurostat)
    - Margaret McLoughlin (Central Statistical Office, Ireland)
    - Jolanta Stefanska, Shahbaz Khan (IMF)

6.   The following issues were raised in the panel discussion:

- Governance is very important, and there are various approaches. All of them try to involve the senior management, subject-mater statisticians and users.
- There should be a focus on outputs and improving timeliness, for example by setting strong deadlines and limiting the number of projects undertaken simultaneously. While some participants believed that limiting projects to 3-6 months would be reasonable, other argued that may be too short for some complex statistical activities.
- Lengthy discussions should be avoided where possible, but may be necessary in some cases to convince the subject matter area before the budget is released. A particular subject-matter expertise may result in requests that particular processes are implemented in a specific way.
- Reuse of components developed by others was suggested as a way for more efficient IT governance.
- IT Infrastructure Library (ITIL) was referred to by almost all presenters. It should be discussed at a future MSIS meeting.

7.   Outsourcing was of great interest during the panel discussion:

- Outsourcing may improve efficiency, but also add to complexity. It requires more attention to be paid to accurately specifying requirements at the outset, gathering users' needs, and testing on delivery.
- Outsourcing may have different issues for international organizations, large and small countries. A selective approach appears to be appropriate for small countries.
- There may be also political issues related to outsourcing, in particular when sensitive systems and/or tasks are concerned. The public perception of data security and confidentiality plays a great role.
- There are different issues with outsourcing operations and outsourcing infrastructure and applications development

8.   In concluding the discussion on this topic, the participants emphasized that statistics have very specific and complex requirements on information systems.

Topic (ii): Statistical information systems architecture
Discussant: Karen Doherty (Statistics Canada)
Documentation: Invited papers by Canada, Italy and Eurostat; supporting papers by New Zealand / UNECE, Republic of Korea, Slovakia, Latvia and Mexico.

9.   The discussion on this topic focused on the role of metadata in statistical information systems, as well as on standardisation issues.

10.   The following points were made during the discussion on approaches to metadata architecture:
   • With a growing central role of metadata in the management of statistical processes, there is a growing need for collecting and updating metadata. This requires more distributed tasks and responsibilities with respect to metadata maintenance. Experiences in using wiki for this purpose were discussed.
   • User acceptance of wiki systems was achieved by focusing on user friendliness and improved metadata quality. Wiki also offer exciting collaboration possibilities, which help their acceptance by staff.
   • Quality is an important issue, and in this respect it is necessary to maintain links between data and metadata. It is, however, meaningful to provide access to methodological metadata (classifications, definitions) in the metadata repository before a survey is finished.
   • The term metadata has a very broad meaning, and it was suggested to distinguish between technical and statistical metadata.
   • UML was suggested as a useful tool for modelling metadata related processes.
   • Statistical offices have to manage changes and historical metadata. Therefore, the period of validity should be attached to metadata items.

11.   The meeting also considered experience with creating and implementing a framework for SDMX data exchange.
   • SDMX implementation requires that metadata for Data Structure Definitions (DSDs) are mapped with metadata actually stored in databases of the statistical offices.
   • DSDs are currently available for download on web services of international organizations – recipients of data from national statistical offices.
   • The target solution is to develop a centralised registry/hub that would let users to access all data and metadata from a single point. An intermediary measure would be to make data and metadata accessible through a single web service linked to distributed databases. Technically, an integrated platform is preferable.
   • In a long-term perspective, not only dissemination in the SDMX format, but also SDMX data and metadata import should be addressed.

12.   The participants also discussed the Generic Statistical Business Process Model. The model provides a common language for describing processes from the initial to the final phases of statistical production. The model is on the agenda of the Work Session on Statistical Metadata (METIS 2008) held back-to-back with the MSIS meting. The discussion pointed out that the issues of metadata, registries and quality are common to all processes.

13.   The participants also followed presentations on a portal service – one stop access to statistical information across many organizations. Another presentation suggested a use of a fuzzy SQL that provides flexibility for the inclusion of marginal records that almost meet the criteria. Queries may be based on linguistic expressions (e.g. "small areas with high unemployment").

14.	The final discussion identified common trends in statistical information systems:
- Standardization efforts related to the infrastructure;
- Improving processes;
- Minimizing resource requirements;
- Generalizing statistical systems;
- Moves towards service –oriented architecture.

15.	The following points were raised in the general discussion:
- Statistical offices have a desire to share components. However, this would require using some common building blocks, thus making a case for further standardization. The Generic Statistical Business Process Model offers a possibility, and there should be cooperation between the METIS and MSIS groups in this respect.
- There is a case for sharing components not only between offices, but also between different systems within an office. One of the challenges is the identification of tools that can be re-used for demands formulated by subject-matter areas. There are often still specific tools in addition to corporate tool sets.
- The need for bridging the Generic Statistical Business Process Model with the service oriented approach was stressed. Statistical processes are not always sequential, as suggested by the model. From the conceptual viewpoint there can be a clear separation between the design time and execution time.
- There is a tendency towards more simplified systems; fewer, but more functional systems. However, there are often separate systems for social and economic statistics. The discussion showed that this is mostly related to different business owners on the social and economic side, and a number of specific differences. For example it is usually easier to produce outputs for social statistics using standard statistical packages. However, there was a feeling that it is possible to consolidate specifications and create unified systems.
- The use of a unified platform and generalized systems would help to overcome continuity issues caused by the staff mobility.
- While wiki are used for collaboration inside some statistical offices, there is not much use of such tools for communicating with users of official statistics. The needs for maintenance and moderation are challenges that usually prevent the use of wiki on the public side.

Topic (iii): Exchange/sharing/re-use of components, common models among statistical offices
Discussant: Marton Vucsan (Statistics Netherlands)
Documentation: Invited papers by Norway, United Kingdom, UNIDO and Eurostat.

16.	There is very limited sharing of components among statistical offices. There are limitations of a technical nature, such as different infrastructures, as well as linguistic and licensing issues. The participants considered different models of sharing and related legal, technical and logistics issues.

17.	The example of the Statistics and Open Standards (SOS) Group was discussed. This has grown from a Nordic contact group, and has become a useful network for informal exchange of ideas and experiences, based on common interests. The components and functions are shared among participating offices despite the lack of a common business process model. The SOS Group is hierarchically organized with a Steering Group and technical subgroups, but the cooperation does not follow strict formal procedures. The cooperation programme has included:
- PC-Axis dissemination tools;
- Cristal (phased out);
- SuperStar (common negotiations with the supplier);
- Classification database based on the Neuchatel model;
- Standards (CoSSI, XML 4DR);
- Development of a common business process model.

18.   Experiences using software developed in the 'R' object-oriented statistical environment were discussed. This open source environment contains a wide range of applications, covering many different phases of the statistical business process. 'R' provides a framework for extensions by creating packages containing R code as well as code written in other programming languages (C, FORTRAN, C++, Java).
- 'R' can be made easy to use by simplified interfaces for routine functions.
- One office reported on moving successfully from SAS to 'R', though another said that the costs of retraining would be too high.
- There were good experiences with support from the open source community, particularly when reporting bugs.
- Some issues were experienced with respect to large data sets, though the latest release seems to be better.
- 'R' might be useful for technical assistance projects, for example in developing countries.

19.   The project 'XML for Data Interoperability in Statistics' (X-DIS) with the aim of improving transmission, interpretation and understanding of statistical data and metadata, was also discussed at the meeting. To ensure the long-term sustainability of the tools developed under X-DIS, Eurostat has set up a special work area on OSS (Open Source Software) under X-DIS.
- X-DIS sharing is available through CIRCA (http://circa.europa.eu/dsis), via the Interest Group 'OSS and Statistics'. There is free public read access for this Interest Group.
- The software is shared under the European Union Public License, which is a modified version of the General Public License (GPL).

20.   The following points were raised in the discussion:
- Cooperation can be undertaken in different ways. There can be a simple sharing of experiences. Cooperation can also extend to harmonization, standardization, the use of common models and the actual sharing of components. To be successful, such cases often need a commitment at the senior management level of the parties involved.
- There may be legal impediments to software sharing. These relate to ownership, sustainable support and financial implications, and it would be useful to discuss them at a future MSIS meeting. Licensing should be clarified in order to avoid grey areas of de-facto sharing without a proper license and user support.
- Two sharing models discussed at the meeting were:
    o Software developed by one country that serves as a leading country for the software in question;
    o Creating a community that would facilitate sharing of a set of software and components and the related multilateral relationships.
- Open Source Software (OSS) is a good approach for public organizations, providing minimum liability but a high level of trust. However, there is no commitment to further development. OSS is not subject to public procurement procedures, and often has very good documentation.
- The PC-Axis model looks useful and pragmatic. It takes the form of a consortium with financial and in-kind participation of members.

21.   The participants also discussed possible actions to be taken in order to facilitate the sharing of software and components between participating countries. The following proposals were made:
- Create an IT development community amongst NSI/ISI(s) interested in making available statistical services/products;
- Establish a governance agreement which comprises a sustainable development and support model for any service made available to the community;
- Community members should establish a common development standard.

Eurostat plans to launch a project to promote cooperation, and suggested creating a Task Force that would prepare specifications for future work.

Topic (iv): User perspective
Discussant: Lars Thygesen (OECD)
Documentation: Invited papers by Netherlands, Canada and Sweden/OECD.

22.   This topic focused on the way in which the information systems put in place by statistical offices interact with various audiences, in particular with respect to quality, interpretability and efficiency. Two primary types of audiences were considered: (i) respondents and (ii) users of official statistics. This means that the discussion mostly concerned systems involved in data collection and data dissemination.

23.   The following points were made with respect to systems for data collection:
- Internet census data collection provides the following benefits:
    o It is efficient for countries and areas with a low population density.
    o Quality is improved; according to experience, edit failure rates were significantly lower for on-line responses.
    o Reduced postage costs, reduced staff input (enumerators and keying in the responses).
- The Public/Private Key Infrastructure (PKI) can be installed, as shown in an example presented, and can be implemented in a way that is transparent to the users. This is particularly useful with respect to respondents that reply to statistical questionnaires only occasionally, and for whom the standard procedure for obtaining the keys would be dissuasive.
- The following should be considered when preparing on-line questionnaires:
    o The questionnaire should be tested for a wide range of browsers.
    o The questionnaires need to be easy-to-use, the respondent should be aware of the progress through the questionnaire, and a version for visually impaired respondents may be needed.
    o Experience shows that helpline operators need a specific layout of the questionnaire.
    o There should be a good cooperation with the network operators, and a backup plan should be put in place. Bandwidth should be managed, and access may need to be limited on the enumeration day – according to experiences, users understand these limitations.

24.   The following points were made concerning dissemination systems:
- Google Earth provides an attractive tool, because it combines statistics with other popular information. In this way statistics add value for users who are already used to Google Earth and have downloaded the plug-ins.
- A dual approach was suggested, with Google Earth used to target more advanced users and Goole Maps for those who will to have just a quick look without downloading and installing any plug-ins.
- Generalized data dissemination systems provide not only one-stop access and a common technological platform, but also a motivation for harmonizing the statistical output (code lists and value lists, standardized dimensions, units, etc.). The issue of standardizing units requires agreement across the organization:
    o A radical solution is to restructure all datasets;
    o A pragmatic solution is to keep datasets, but repair them. It was noted that some data owners even found the pragmatic solution too radical.

25.   The general discussion included the following points:
- There is a need to keep the contents fresh in dissemination systems, to keep the interest of users.
- For visualization purposes, different data sets need different scales so that variations and movements can be easily seen.
- SDMX provides a useful standard for communication between databases and visualization tools.
- There is a mode effect between web and other forms of data collection, particularly in the level of detail given in free text responses.
- A group was proposed to consider IT issues for censuses. A wiki may also be useful for exchanging ideas and experiences on this topic.

* * * * *