

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

**EUROPEAN COMMISSION
EUROSTAT**

Joint UNECE/Eurostat Work Session on Electronic Raw Data Reporting (ERDR 2006)
Geneva, 6-8 November 2006

REPORT

1. The Joint UNECE/Eurostat Work Session on Electronic Raw Data Reporting (ERDR 2006) was held in Geneva from 6 to 8 November 2006. Delegates from the following countries attended: Australia, Austria, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Lithuania, Luxembourg, Netherlands, Norway, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom and United States. Experts from XBRL Europe and Agilis SA attended the meeting at the invitation of the UNECE secretariat.
2. The Work Session was organized as a joint event of the Eurostat-sponsored Group on Collection of Raw Data (CoRD) and the UNECE-sponsored Task Force on Electronic Raw Data Reporting for Primary Data Collection (ERDR).
3. The agenda of the meeting consisted of the following substantive topics:
 - (i) XBRL and business reporting;
 - (ii) Electronic reporting of census data and other national experiences;
 - (iii) Open source software for electronic raw data reporting;
 - (iv) Authentication, security and privacy issues.
4. In their opening addresses the representatives of the UNECE and Eurostat explained the mandates of their organizations in the area of electronic reporting of data for primary data collection. As the mandates are similar, the two organizations decided to join forces to pursue the activities in the course of the next two years.
5. Mr. Leonhard Maqua (Eurostat) chaired the meeting.

RECOMMENDED FUTURE WORK

6. The following tasks and deadlines were agreed for the future work of the Task Force:
 - Develop a structure/framework for the information warehouse – through electronic consultation (December 2006);
 - Nominate moderators for chapters/sections of the warehouse (January 2007);
 - Solicit and feed the content, with assistance of other participants (January-October 2007);
 - First progress report to the CES Bureau (structure and organization of work) (February 2007).
7. The participants suggested the following leadership of the Task Force, subject to clearance by their respective national statistical offices:

Chairman:	Mr. Fritz Pfrommer, Germany
Vice-Chairs:	Mr. Johan Lammers, Netherlands
	Mr. Paul Williams, Australia
	Mr. Rune Gløersen, Norway
	Mr. Bertrand Loison, Switzerland

Secretaries: Mr. Juraj Riecan, UNECE
Mr. Leonhard Maqua, Eurostat

8. The next Work Session on Electronic Raw Data Reporting was proposed for November 2007 to review and prepare for final approval the content within the information warehouse on ERDR. Eurostat will organize a meeting of the CoRD group in June 2007, preferably back-to-back with the STNE meeting.

FURTHER INFORMATION

9. The conclusions reached during the discussion of the substantive items of the agenda are contained in the Annex. The papers that served as background for the discussion along with the PowerPoint presentations are available on:

- the UNECE website <http://www.unece.org/stats/documents/2006.11.erdr.htm>;
- the CIRCA service of Eurostat (<http://circa.europa.eu/dsis> -- ERDR Task Force). The CIRCA service will also be used to facilitate the discussion among members of the Task Force after the meeting.

10. The participants adopted the present report before the Work Session adjourned.

Annex

Summary of the main conclusions reached by the participants during the discussion at the November 2006 UNECE/Eurostat Work Session on Electronic Raw Data Reporting

Topic (i): XBRL and business reporting

Documentation: Papers by Luxembourg, Netherlands and XBRL Europe

1. The presentations recalled the history of XBRL (Extensible Business Reporting Language) and recent developments:
 - Accountants developed XBRL at the end of the 1990s.
 - XBRL standard maintenance is a non-profit activity, not linked with any commercial or political interests. The role of XBRL is to provide technical specifications.
 - XBRL is currently being used within some national and international institutions, either as a mandatory or optional standard.

2. Participants discussed the opportunities XBRL provides for reporting business data to national statistical offices. They identified the following issues:
 - Whether XBRL is the future standard in business reporting. The answer depends mainly on businesses, banks and public administration. The general government approach to XBRL also significantly influences the answer to this question. Official statistics will have to draw lessons from this development, but it does not play a lead role.
 - Reporting for statistics has specific requirements that are not always reflected in other reporting flows. Therefore, an effort is needed to adjust financial and other business reporting to statistical needs. The weight of this effort differs from country to country, making XBRL more attractive for some countries, while others may prefer other techniques. The content and opportunities will influence the choice of the statistical office.
 - Small- and medium-sized enterprises (SMEs) represent the most important user group with respect to the number of expected reports (tax returns, annual accounts, etc.). While the benefits may be more for businesses rather than statistical offices, this should still be considered positively from the national viewpoint. It was emphasized that particular data (e.g. exports and imports) needed for statistical offices are already available within SMEs. These are the most likely candidates for participation in business reporting.
 - The notion of “raw data” is relative, and may somehow differ between XBRL and statistics. For example, the XBRL GL (Global Ledger) seems to be “too raw” in some cases. Statistics may need more aggregation, and the content issues have to be addressed.
 - There is a lack of awareness about XBRL. National and international activities should focus more on exchanging XBRL-related experiences, future plans, and the contacts in individual countries. Eurostat has a task force on XBRL that can play a role.
 - In several countries XBRL has a place within global government policies. Countries aim to decrease the administrative burden by, for example, creating common entry points for all data (e.g. Norway, Luxembourg, Netherlands, etc.). While it is not likely that statistical offices will bring the idea of XBRL reporting to the governments, they may participate in the initiative after XBRL gains general acceptance.
 - It is important that the content can be viewed by statisticians, and that they obtain the necessary meta-information. Two-way mapping between metadata and XBRL taxonomies is required. The XBRL taxonomies are hierarchical code lists of terms, attributes, concepts and their interrelationships in different types of business reporting.
 - Support by software vendors is a necessary pre-condition for broad implementation of XBRL, not only within statistics. Such support, usually by international vendors, would have to consider specifics emanating from different national legislations and reporting patterns.

3. The issue of data quality is a central concern for official statistics. The discussion also focussed on how XBRL could influence quality:
 - While XBRL and other similar electronic transmission modes would ensure that there is no loss of quality during transmission, there are still fundamental quality-related issues to be resolved. Improvement can be achieved through content standardization.
 - Regarding data editing by respondents, participants considered that placing a more manual burden on respondents could suppress the benefits offered by XBRL through automated reuse of accounting systems. The exchange of information between the ERDR Task Force and the expert group on data editing might bring a common understanding to future directions.
 - Timeliness is another important quality issue. Experience has shown that responding using electronic reporting is usually more timely than paper reporting.

4. The participants also discussed the development of common taxonomies in collaboration with other ministries and government offices:
 - The information supply chain has a number of participants, and working together with them is more likely to bring tangible results than developing specific taxonomies for statistics.
 - The information needs of statistics, businesses, tax offices and other public services overlap. This favours the creation of common taxonomies.
 - It is important not only to define a single entry point for each type of data, but also for the content and procedures for sharing data among government agencies, so that the data meet all purposes.

5. Participants were informed that the 14th XBRL International Conference is scheduled to take place in Philadelphia, United States, from 4 to 6 December 2006.

Topic (ii): Electronic reporting of census data and other national experiences

Documentation: Papers by Australia, Portugal, Slovakia, Spain and United States

6. The use of the Internet for census data collection can fulfil different roles:
 - In support of interviewers, for providing the information as well as data transmission using portable devices;
 - In support of respondents using traditional paper forms;
 - As an on-line response option.

7. The discussion focussed on the on-line response option.
 - At the initial phase the Internet on-line option represents increased costs for the statistical office. There are some savings in the processing costs, but they are less important in the start-up phase.
 - A portion of the public expects an on-line response option. In some countries, government policies oblige all public services to offer an on-line option. Electronic banking and electronic businesses provide further incentives for pursuing the development of the on-line option.
 - On-line forms logically allow embedding respondent-side editing within the forms. However, mandatory edits may increase the risk of non-response, and it is advisable to proceed cautiously.
 - One suggestion was to make the on-line forms available in parallel with the distribution of the paper forms. This may result in response before the census day, but such time deviations may be acceptable for the sake of an increased response rate.
 - The paper and on-line forms should be designed for the same content. Some participants also considered it useful if the paper and electronic forms are similar – but the approach differs from country to country. In all cases, the statistical offices kept the electronic forms as simple as possible.

- In some cases, duplicate answers were experienced. A probable cause is that respondents wanted to test the electronic option, but they still preferred to use the paper form.
8. The discussion also touched on the issue of pre-filled census questionnaires, when using the on-line response option.
- Pre-filling of on-line forms is similar to pre-filling of traditional paper forms. A precondition for this is availability of population and other relevant registers.
 - Linking of the registers also requires developing supporting IT systems. This is different in each country.
 - Very often countries that are able to pre-fill the forms choose to use register-based censuses. They do not use any forms addressed to the population.
 - A special case is the “rolling” census, when each year a portion of the population is enumerated. In this case, even if the census data collection takes place annually, the pre-filling of forms would not be sufficiently accurate.
9. The discussion also emphasized the necessity for adequate advertising when offering an electronic response option:
- Advertising for the on-line response option should be well planned and should begin well in advance of the census.
 - In some cases, enumerators who are more used to traditional forms of census data collection can play either a persuasive or dissuasive role.
10. Some national experiences in using electronic data reporting for survey data collection were also discussed at the Work Session:
- There may be some technical similarities between electronic raw data reporting for surveys and for censuses, but the purpose of the two types of operations is different. There are, therefore, differences in the relationship with the users, periodicity, repetitiveness, the utility of pre-filled questionnaires etc.
 - The examples discussed demonstrated the use of the Internet for:
 - Collection of data directly from users;
 - Transmission of data from regional/branch offices to the centre.
 - Promotion of on-line data collection contributes significantly to the respondents’ choice and progressively growing cost recovery. It was suggested to customize the promotion according to user groups, and to create incentives for electronic responses.
 - The on-line response option brings some advantages to data quality:
 - Through the built-in checks (edits);
 - Through offering electronic communication with the respondent, so the ex-post verification is more dynamic and more likely to be answered.
 - Discrepancies often occur when information on the company changes (typically: the name of the company). This change is reflected in the business register after certain legal deadlines. On the other hand respondents have a different perception and reflect the change earlier (or later) than the change is recorded in the business register.
 - Requirements for authentication of responses differ in each country. Some require registration and ensure that the response is coming from a designated respondent. Others, which do not consider the risk of abuse to be very important, do not require registration.

Topic (iii): Open source software for electronic raw data reporting

Documentation: Paper by Eurostat

11. Open source software (OSS) should not be confused with freeware or public domain (no copyright) software.

12. The main advantages of OSS are:
 - Interoperability – the code can be compiled for different operating systems, and adjusted to different applications. However integration with proprietary documents and file formats can be problematic.
 - Availability of the source code.
 - Security – no hidden “back doors”.
 - Quality is often higher as the source code is available for checking / peer review. This was the feature emphasized by the IT experts present at the meeting.
 - Low costs – OSS usually has lower costs than typical proprietary software. Often it is free, but not as a rule. Although acquisition costs are low, other costs, such as migration, training and support, must be considered.
 - Stability – users are not forced to migrate to new versions.
 - Independence from the dominant software vendors – even if the vendor ceases to exist, the users can continue development of the software independently. On the other hand, the software producers are not obliged to commit to long-term development.

13. An interest group “Open Source Software and Statistics” is available on CIRCA. Participants were encouraged to use and share the information available there (http://forum.europa.eu.int/irc/dsis/oss/info/data/en/home_page.html). The goal of this website is to become a one-stop source of information on OSS for statistical administrations. It is updated and developed regularly. In the future it will be enhanced with the results of a survey to be conducted in 2007 on the use of OSS in statistical organizations.

14. The **GNU** General Public License (GPL) model is a popular license for OSS. The European Union Public Licence (EUPL), developed by DG Enterprises and Industry, has been approved by the European Commission’s legal service and will soon be formally adopted by the European Commission as the preferable licence scheme to be used by Commission services.

15. The following technical issues related to the use of OSS for electronic data reporting were raised in the discussion:
 - For the purpose of the electronic raw data collection, OSS offers the following advantages:
 - Better acceptance by users (mostly respondents);
 - Possibility of sharing development between administrations; and
 - Customization for particular needs of statistical organizations.
 - Common formats already exist for aggregated data exchange, and Eurostat plans to publish several tools for SDMX under the EUPL license.
 - There is still work to be done on developing formats for raw data reporting.
 - While the OSS is interoperable to a large extent, limitations exist in the programming languages used within statistical organizations (Java, Microsoft .Net, etc.). This could restrict the possibility for the general sharing of the software.

16. Issues of standardization were considered crucial for further development of statistical OSS. In-depth discussion focussed on:
 - Standardization of formats for raw data by national statistical offices. They have better knowledge of the issues than international organizations.
 - Metadata harmonization is another important condition for implementing OSS in statistical data collection.
 - Agreed common formats and metadata standards can only be developed effectively with input from a wide community.
 - SDMX will bring benefits when it becomes possible to report/exchange data and metadata with more than just the current group of sponsors.

17. General aspects also influence the possibilities offered by OSS:

- Countries need to localize their software systems by implementing them in their official language(s). This may affect the ability to share OSS between organizations.
- OSS can discharge the statistical offices which have developed it from giving guarantees and extensive support to end-users, as users can fix the bugs in their own environment. On the other hand the investment would not be returned through commercializing the software.
- While OSS looks promising for public administration, and it has been discussed already for several years and there is still not much concrete development.

Topic (iv): Authentication, security and privacy issues

Documentation: Papers by Estonia, Finland and Agilis SA

18. While the Internet offers a valuable tool for electronic data exchange, it raises several security concerns:

- The Internet is vulnerable to a variety of threats. Sensitive information may be fraudulently obtained by listening to communication, re-routing web pages, etc.
- There are several methods to decrease the risk of attack, but there is no perfect solution:
 - Forms authentication is vulnerable to brute force dictionary attacks.
 - Secure Socket Layer (SSL) technology is vulnerable to 'Man in the Middle' attacks.
 - Password protection may be ineffective due to reuse or weak passwords, or writing the password down, making it accessible to others.
- Many organizations have not developed effective security strategies.
- Important aspects of security to consider are: confidentiality, integrity, availability, authentication, and accountability.

19. In recent years, the XML syntax is replacing the UN/EDIFACT syntax. XML is not just a syntax, but a family of technologies that allow the establishment of an environment of data exchange. This allows content and structure standardization, and offers new possibilities for security:

- It is possible to encrypt only sensitive parts (statistical data) of the XML message. Administrative information needed for routing, as well as metadata should be accessible within the non-encrypted part of the XML message.
- Java libraries are available to developers for implementation of XML security. Information on these will be made available by the CIRCA interest group on the OSS and Statistics (see Topic (iii)).
- Digital signatures may be used for authentication of XML message, if needed. Public domain software is available for this purpose.
- Security strategies must take into account that XML is a language for machines rather than humans.

20. Participants discussed the use of personal identification to authenticate users completing electronic business surveys:

- Different solutions are dependent on the PKI infrastructure of the countries.
- The purpose is to identify authorized users that may access sensitive economic data. This is applicable when respondents have access to past data, as is the case in some countries.
- Machine-readable identification cards, or Internet bank cards can be used for personal identification. However, this would not work in countries that do not have machine-readable ID cards. Moreover, in some countries the use of Internet bank cards could be badly perceived.
- Additional hardware is needed, such as a standard card reader.

21. A complex solution that provided an advanced level of security for a national statistical office includes the following features:

- Role-based access control instead of a simple authentication;

- Credentials mailed through the postal system;
 - Extensive logging of online activity;
 - Encryption.
22. Participants considered to what extent it is possible to share solutions for security:
- Security strategies and issues must be kept confidential in order not to increase security risks.
 - Knowledge about attacks that have occurred can be shared. Hammering attacks and viruses that attempt to read keystrokes entered into online systems were quoted as examples.
 - Risk management is needed rather than risk avoidance, because there is no absolutely reliable way to avoid the risk.
 - Statistical offices recruit external companies to conduct security audits, but the findings are confidential. Some general issues emerging from these exercises may be shared, but without going into detail.

Future work

Documentation: Terms of Reference and Work Plan for the ERDR Task Force

23. The following points were made in the discussion on the future work
- Censuses are a specific issue within this theme. They also cover much more than ERDR, such as use and quality of registers, etc. In this respect, it might be better to bring the discussion to a specialized forum, such as the UNECE expert group on population and housing censuses. The information on censuses and ERDR should be collected on CIRCA.
 - It is important to maintain a forum about what is going on in other countries, who to contact for further information. For this purpose, everybody has to contribute. The Task Force should aim to put together the information on practical solutions, with examples of implementations.
 - According to the Terms of Reference, the main purpose of the Task Force is to document the knowledge and experiences. This suggests creating an information warehouse (on CIRCA or somewhere on the Internet), where people would be able to find information on topics related to the electronic collection of data that interest them.
 - A lot of information was collected within the CoRD Group of Eurostat. It would be useful to put together information on the state-of-the-art in countries, with links to actual demo systems, so that everything can be found in one place.
 - There are other international events related to electronic raw data reporting. The information on conferences should be put on a special newsgroup on CIRCA. The ICES III Conference to be held in Montreal in 2007 was mentioned as an example.
24. Some topics of interest for future work were mentioned:
- Architecture for data collection is not explicitly mentioned in the list of topics, but is of interest to participants.
 - Metadata-driven ERDR systems. This would require in-depth work on metadata that is undertaken by the METIS group. UNECE and Eurostat should ensure the cooperation between the two groups.
 - The use of the Internet for household surveys is the theme that may deserve further discussion. In some countries, there is more scepticism about the usability of ERDR for household statistics as opposed to the business statistics.
 - Eurostat's Task Force pursues the work on XBRL anyway, and the results might be brought to the attention of the ERDR Task Force. There is also an interest in other XML-based standards, such as X-forms. X-forms are not widely used probably owing to a lack of awareness.
 - More clear documentation on IQML would be needed to have a better understanding of how to use it. Participants stressed that there may be some work remaining in this direction.

25. OSS and sharing software solutions.
- Some participants expressed scepticism about the re-usability of the software among agencies. Others have witnessed a convergence in solutions (for example IQML), and were slightly more optimistic. There was general agreement that it is interesting and important to know what software other offices use.
 - A concrete step would be to create an inventory of available software, with the description that is needed for those who may be interested in the software.
- 26 The following tasks and deadlines were agreed upon for the future work of the Task Force:
- Develop a structure/framework for the information warehouse – through electronic consultation (December 2006)
 - Nominate moderators for chapters/sections of the warehouse (January 2007)
 - Solicit and feed the content, with assistance from other participants (January-October 2007)
 - First progress report to the CES Bureau (structure and organization of work) (February 2007)
27. The participants suggested the following leadership of the Task Force, subject to clearance by their respective national statistical offices:
- Chairman: Mr. Fritz Pfrommer, Germany
- Vice-Chairs: Mr. Johan Lammers, Netherlands
Mr. Paul Williams, Australia
Mr. Rune Gløersen, Norway
Mr. Bertrand Loison, Switzerland
- Secretaries: Mr. Juraj Riecan, UNECE
Mr. Leonhard Maqua, Eurostat
28. The meeting recommended that the next Work Session on Electronic Raw Data Reporting be held in November 2007. It is expected to review and prepare for final approval the content within the information warehouse on ERDR. Eurostat will organize a meeting of the CoRD group in June 2007, preferably back-to-back with the STNE meeting.

* * * * *