

Distr.
GENERAL

CRP.1
18 May 2010

ENGLISH ONLY

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

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC CO-OPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Meeting on the Management of Statistical Information Systems (MSIS 2010)
(Daejeon, Republic of Korea, 26-29 April 2010)

REPORT OF THE MEETING ON THE MANAGEMENT OF STATISTICAL INFORMATION SYSTEMS

Prepared by the UNECE secretariat

1. The Joint UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS) was held in Daejeon, Republic of Korea, from 26 to 29 April 2010. Participants from the following countries attended the meeting: Australia, Belarus, Bhutan, Cambodia, Canada, China, Croatia, Estonia, France, India, Indonesia, Ireland, Israel, Italy, Malaysia, Mongolia, Netherlands, Norway, Poland, Republic of Korea, Russian Federation, Sri Lanka, Sweden, Switzerland, Thailand and United Kingdom. The European Commission was represented by Eurostat. Representatives from the following international organizations also attended: United Nations Economic and Social Commission for Asia and the Pacific (ESCAP), Institute for Statistics of United Nations Educational, Scientific and Cultural Organization (UNESCO), United Nations Industrial Development Organization (UNIDO), World Health Organization (WHO), Organisation for Economic Co-operation and Development (OECD), International Monetary Fund (IMF), and Bank for International Settlements (BIS).
2. The agenda of the meeting (WP.1) consisted of the following substantive topics:
 - (i) Developing common high-level architectures
 - (ii) Software sharing and shared maintenance
 - (iii) Issues for Asian statistical organizations and ways to increase international cooperation
 - (iv) Innovation and related issues including census systems
3. Mr. Rune Gløersen (Norway) was elected Chairman of the meeting. The preparation of the substantive work was organized by: Ms. Karen Doherty (Canada) for Topic (i); Mr. Marton Vuksan (Netherlands) for Topic (ii); Ms. Jung-Im Ahn (Republic of Korea) and Mr. Ilpo Survo (ESCAP) for Topic (iii) and Ms. Jolanta Stefanska (IMF) and Mr. Trevor Fletcher (OECD) for Topic (iv).
4. Mr. Seung Wou Shin, Director-General of the Statistical Information Bureau of Statistics Korea opened the meeting and welcomed participants. He noted the advancement of statistical comparability and information systems, and stressed that the meeting would be very useful for national statistical organisations from the Asian region.
5. In his introductory remarks, the Chairman noted increasing interest in the work of MSIS. It enables statistical organisations to take stock of each other's achievements and cooperate better at the international

level in fields of common interest, such as architecture. He introduced the members of an ad hoc group to solicit ideas and prepare proposals on future work, to be presented for discussion on the final day. This group consisted of the Chairman himself, Ms. Branka Cimermanovic (Croatia), Mr. Carlo Vaccari (Italy), Mr. Simon Field (United Kingdom), Mr. Valentin Todorov (UNIDO), and Mr. Marton Vuksan (Netherlands). Participants were encouraged to contact the members of this group with ideas for the future work priorities relating to the management of statistical information systems.

6. The future work group prepared the following proposals for topics to be discussed at the 2011 MSIS meeting, which were endorsed by the delegates:

- Topic (i): Architectures, models and standards
 - Definitions and scope of architecture and high-level models (business process models, information models etc)
 - Managing reference metadata (Data Structure Definitions etc.)
 - Metadata driven systems, challenges and opportunities
 - Data Documentation Initiative (DDI) and Statistical Data and Metadata eXchange (SDMX), coherence, usage etc.
 - Metadata vocabulary, statistical ontology, semantic web and linked data
- Topic (ii): From local to corporate perspective
 - Technology
 - Service-oriented architecture
 - Management
 - Business continuity
 - Data and metadata management
 - Skills, capability
- Topic (iii): Innovation and related issues
 - Demonstration of bleeding edge software solutions
 - Best practice examples
 - Geographic data
 - Cloud computing, Software as a service (SAAS)
- Topic (iv): International cooperation/collaboration
 - Reference models for international cooperation/developments
 - Examples
 - Benchmarking
 - Framework
 - Examples
 - Report and work program for the Sharing Advisory Board

7. In his summary the Chairman said that the meeting had been very fruitful, providing a useful opportunity to share experiences and challenges. As a result, participants have new ideas and new contacts, and many had expressed relief not to be the only ones facing particular problems. He thanked the MSIS Steering Group and the UNECE secretariat for their work preparing the meeting.

8. The Chairman and participants expressed particular thanks to the Korean Statistical Office for the very high quality of the local organisation, the excellent facilities and their generous hospitality.

9. The conclusions reached during the discussion of the substantive items of the agenda are contained in the Annex. Presentations and all background documents for the meeting are available on the website of the UNECE Statistical Division (<http://www.unece.org/stats/documents/2010.04.msis.htm>).

10. The participants adopted the draft report before the Meeting adjourned.

ANNEX

SUMMARY OF THE MAIN CONCLUSIONS REACHED DURING THE MEETING ON THE MANAGEMENT OF STATISTICAL INFORMATION SYSTEMS (MSIS 2010)

Topic (i): Developing common high-level architectures (e.g. business, information and IT architectures)

Session Organizer/Discussant: Karen Doherty (Statistics Canada)

Papers by: Canada, Australia, Ireland, Malaysia, China, Republic of Korea, Israel, Sweden, Italy/Eurostat, and Belarus.

1. The papers presented under this topic covered a range of topics from overall business architecture to the architecture and design of specific applications and systems. A common theme was that statistical organisations are looking for cohesive systems, and developing new and efficient information technology (IT) strategies.

2. Statistics Canada has found that respondents are harder to reach through traditional methods so programme areas are finding it increasingly difficult to maintain the quality of their data products and is struggling to deal with a series of budget cuts. To address these issues Statistics Canada has embarked on a complete renewal of the agency's corporate business architecture. This will optimize decision-making at the corporate level; change the statistical process to make it meta-data driven; maximize the use of corporate services; maximize re-use and strengthen information management. Electronic data collection is the preferred collection mode. The business process model is based on the Generic Statistical Business Process Model (GSBPM). The object is to manage quality and to decide what is "good enough" and not invest where it isn't necessary. IT services currently comprise around 1000 staff, split between the IT branch and business areas. These staff will be centralized in the IT branch, and reduced by 10% over the coming years. There is a need for strong governance and support from senior management, clear funding models and good communications. Questions concerned the availability of additional funding, how to manage prioritisation and the loss of flexibility in business areas, metadata management, division of tasks, recruitment and retention of IT staff, and the time required for planning.

3. The Australian Bureau of Statistics presented their collaborative approach to agile statistical processing architecture. They are developing a collaborative approach based on international standards, and have adopted the GSBPM. Future data processing will incorporate the Data Documentation Initiative (DDI) standards for micro-data and Statistical Data and Metadata eXchange (SDMX) standards for aggregate data. Under this approach the focus is on mutual value propositions and the realisation that it can be desirable to have multiple solutions. The new Information Management Transformation Project focuses on metadata management throughout the statistical business process, with an emphasis on metadata standards and the use of paradata to drive statistical production. Australia supports international initiatives such as the Sharing Advisory Board and the Common Reference Architecture (CORA) project. Applications architecture is seen as belonging to business and it should reflect business needs.

4. The Irish Central Statistics Office uses the GSBPM as a communications tool to explain the rationale of their IT strategy and to support good governance. Inputs taken into account in preparing the strategy included user concerns and priorities, a corporate strategic review in 2009, lessons learned from the previous data management system project, and organizational culture. The GSBPM is a way of influencing corporate culture through planning, production and evaluation, giving greater visibility to quality management, costs and the value of the IT contribution. It is more than just an IT framework; other uses at the corporate level form a valuable part of the GSBPM. Changing the organisational culture is a gradual process, aided by improved communications, top management commitment and a greater focus on standards and quality management. Around 70 out of a total of 750 staff work in IT, and apart from census processing, there is currently no outsourcing.

5. The representative of the Department of Statistics of Malaysia presented their National Enterprise-Wide Statistical System (NEWSS). NEWSS is a by-product from the ICT (Information and Communication Technologies) projects identified in DOSM's ICT Strategic Planning (ISP) - a process for developing a strategy and plans for aligning enterprise level information systems with the business strategies of DOSM. It is an integrated statistical system framework to efficiently monitor and manage surveys and censuses. It extends the dissemination of statistical information and improves the timeliness and reliability of statistical information systems. Complementing the ISSF on assisting the user in performing the effective and efficient process is the ISS (information support systems). ISS enrichment the capability of governance, digital mapping analysis, advance analysis and create the knowledge sharing environment within the department. Phase 1 of the system has taken 18 months to develop, under Ninth Malaysian Plan's budget.

6 The National Bureau of Statistics of China presented their study about statistical information and communication technology for a national statistical organisation. It discussed the role of architecture, the principles for building a system like the one they have recently developed, and the need for international cooperation in this field. The Chinese statistical information and communication technology system covers all stages of statistics production, including questionnaire design and distribution, data collection, data processing and dissemination. It comprises a network, a software platform and a set of databases. Since the production processes of official statistics are standard, it would be useful to develop generic software for different parts of the statistical production processes. Different departments have their own working databases, but there is one central database for aggregate data. China intends to use some aspects of SDMX, and follow the basic principles of that standard, but have not yet gone into detail on implementation.

7. Statistics Korea presented their enterprise architecture. Previous increases in IT resources and costs have not remedied problems, due to the lack of effective governance mechanisms. There were many problems related to surveys, so a new approach was needed to support the production of high quality statistics and to build information systems more efficiently. The strategy aimed at providing achievement-based support for business objectives; to standardize and reuse application systems; to standardize data and metadata management; and to reinforce the security system. There is decentralized statistical collection throughout different government agencies, but statistical business processes and systems in these agencies will be reviewed, and a detailed plan for integration will be prepared. Korean statistical business processes have been mapped to the GSBPM, and gaps relating to the "specify needs" and "evaluate" phases will be addressed by redeploying staff. The integration of administrative data is supported by the Ministry of Public Administration and Security.

8. The representative of the Israeli Central Bureau of Statistics presented their work toward a generic system, as they currently have over 190 statistical information systems. They also plan to change their strategy on information systems development and to decrease the development time for new systems. Generic survey management tools significantly shorten the time needed to create a new survey, however, creating a specific questionnaire, especially in different languages, is time consuming. Return on investment was improved (if a computerized questionnaire is made – one month is sufficient to start the survey); the tool and know-how bought for the administrative file will serve the entire organization; and it reduced the number of information systems in the organization. There are, however, increased risks associated with this approach, notably that a single system failure can have much wider effects. Plans for software sharing are being developed, for example code for "Blaise" supporting Arabic, Russian and Hebrew languages can be shared if anyone is interested.

9. Statistics Sweden presented their event-driven service-oriented architecture (SOA) for data collection and editing, TRITON. They have challenges similar to those mentioned in previous presentations, and the event-driven SOA approach met the criteria for performance and reliability. The new architecture had to be metadata driven and meet quality requirements such as stability and security. The focus is on the business objects to support business processes. A "Generic Statistical Business Information Model" was proposed to provide a common view of the relationships between objects. The architecture is applicable to data collection via surveys and from administrative sources. Mapping to the DDI 3.0 standard was proposed as a future task.

10. Italy presented a joint project with Eurostat on SDMX architecture for data sharing and interoperability. SDMX can help improve quality and efficiency in the exchange and dissemination of data and metadata; reduce national reporting burden; reduce costs through the re-use of the software; facilitate and standardize the use of new technologies such as XML and Web services. This project aims to design an SDMX service infrastructure for national statistical organizations, developing related building blocks and providing financial support through the SDMX “ESSnet” project. This architecture will support the Census Hub application and other Eurostat projects. It will facilitate SDMX implementation within the PC-AXIS suite of dissemination tools and stimulate an SDMX community of developers. The SDMX registry is used in data repository architecture, but not for the census hub application. The software is open source, and can be downloaded by anyone from the SDMX website.

11. The development of modern IT in statistical operations was presented by the representative of Belarus. This new system will be an essential element of e-government and will improve the flow of information between citizens and government. This project will create a new integrated statistical information resource using modern software and hardware. The first step is the construction of computing networking infrastructure, linking national and regional data centres. The statistical information model contains formalized descriptions of statistical data and their interrelations. The new system provides a number of functionalities for data collection and dissemination, including information security management and links to geographical information systems. The new system will be introduced gradually rather than as a sudden change.

12. The session organizer identified the following four common themes from the presentations:

- The importance of management support;
- Sharing of software and ideas;
- Re-use of components;
- Using metadata to drive the process.

13. Points raised in the discussion included:

- How to measure productivity gains from the introduction of new systems and architectures. This is difficult when other changes are taking place, and often the main gains are facilitated by new systems rather than being a direct result of their introduction, for example new outputs, reduced burden on respondents and data quality improvements.
- Organizational and cultural issues are important, as they have a strong impact on the chances of success of IT developments. Organizational change is often needed to realize the benefits of system changes, and can often meet resistance from staff.
- Changes should be driven by the business, and the role of business architects is important, a partnership approach between IT and business areas is essential to delivering effective change.
- There are often bottlenecks related to resources in business areas, and there is a strong need to get buy-in from methodologists, encouraging them to think in terms of standard rather than individual solutions.
- The current wave of developments in architecture is more concept and model driven than previously, therefore common concepts and models (for example a common way of describing statistical objects) are very important.
- Standardized and centralized systems can increase the impact of a single system failure. Business continuity planning and risk management therefore become more important at the corporate level.
- External factors may affect the design of systems and architectures, and can not always be planned for.

Topic (ii): Software sharing and shared maintenance

Session Organizer/Discussant: Marton Vučan (Statistics Netherlands)

Papers by: Republic of Korea, UNIDO, Norway/Sweden, OECD, Italy and Netherlands

14. Statistics Korea presented their spatial statistical information services. The Statistical Geographic Information Service (SGIS) improves the user-friendliness of data through visualization tools. The Census

Individual Spatial Database provides position information for residences and establishments and information about boundaries to support censuses of population, housing and establishments. Systems focused more to users include information on neighbourhoods to help users find those that meet certain criteria, dynamic population pyramids and maps of surname distributions. The levels of detail available for external users are constrained by legal requirements for data confidentiality.

15. The representative of UNIDO presented R in the Statistical Office. R provides an open-source system and programming language for statistical computation and graphics. The core application can be extended using “packages”, which are freely available, covering a wide range of statistical functionalities. It can therefore work with many common data formats, including XML and SDMX. The Sweave application combines R with LaTeX to produce reports containing text and high quality graphics. There is a large R user community, and it is becoming very popular with students and younger consultants. It can operate alongside SAS, and provides similar functionality.

16. The Session Organizer provided an overview of the paper submitted by Cambodia on Data documentation, dissemination and exchange at their NIS and referred participants to the paper on the website.

17. The representative of Norway gave an update on the governance arrangements for PC-Axis. The Nordic countries are considering a “community source” model as the basis for sharing responsibility for the future development and maintenance of the PC-Axis software family. This would involve a core group of users contributing components, whilst other users would contribute financially. In the longer term, a more open-source approach might be possible, but the financial implications of this need to be resolved. The objective is to start working according to a cooperative model by 1 January 2011, but the exact details have not yet been agreed.

18. The OECD representative presented an update on their experiences of sharing statistical software. OECD has been sharing its software since 2007 following requests from other organizations. Sharing was initially informal, but as requests increased, more formal arrangements have become necessary. A positive impact is that their software includes SDMX functionalities, so sharing could encourage the uptake of SDMX by other organizations. The increase in demand has created new legal and administrative challenges, as resources are required to support software sharing, but finding a suitable mechanism to recover costs is not easy.

19. The representative of Italy provided an update on the Common Reference Architecture (CORA) project, which is run under the ESSNet framework with European Union funding. The project contains five work packages: project management; requirements collection and state-of-the-art; technical architecture; organizational architecture; and project dissemination. The aim is to develop a high-level architecture based on the GSBPM. The work of CORA is closely linked to that of the Sharing Advisory Board (SAB). The SAB will provide a way to disseminate some of the information gathered by CORA via the MSIS wiki, and will provide longer-term continuity for this work. A follow-up ESSNet project is proposed for next year, focusing on a common reference environment, and possibly developing software to operationalize the GSBPM. All CORA outputs will be available under a Creative Commons license.

20. The representative of The Netherlands presented the work of the Sharing Advisory Board, including governance arrangements and membership. He outlined the objectives to provide support to software sharing and joint development initiatives. Progress against the 2009-10 programme of work was outlined, and a draft programme of work for 2010-11 was presented for discussion. The Sharing Advisory Board pages on the MSIS Wiki were briefly presented, and it was stressed that the full potential of the resources available will only be realized if the statistical IT community actively contributes to their development and maintenance via the wiki. The work of the Sharing Advisory Board was seen as important, and the future work programme was seen as useful and realistic.

21. The discussion focused on the role of R in statistical organizations. Whilst some organizations see it mainly as a tool for research and development activities, others see it as capable of supporting mainstream statistical production processes. Several organizations use R in combination with SAS.

22. The session organizer led a brainstorming session on possible future directions and outputs for the Sharing Advisory Board work programme, which resulted in the following suggestions:

- Right granularity in components;
- Hands on approach about software sharing;
- Consider language needs in application design;
- Sharing problems and how to avoid them;
- Expanding GSBPM with GSBIM;
- Good communication with PC-Axis community;
- Case studies of sharing;
- Establish what is needed to make software sharable;
- Create documentation / tutorials as well as code;
- Share plans / strategies;
- Shared buying;
- Sharing information about cost of software;
- Information about software / evaluations;
- Documentation about expertise – who to ask;
- Sharing feedback from those who use software;
- Commercial companies to post software information on wiki;
- Share experience in dealing with commercial vendors;
- Do more to share requirements with commercial vendors as a group;
- Clear support mechanisms;
- Share specifications.

Topic (iii): Issues for Asian statistical organizations and ways to increase international cooperation

Session Organizers/Discussants: Jung-Im Ahn (Statistics Korea) and Ilpo Survo (ESCAP)

Papers by: Cambodia, India, Bhutan, Republic of Korea, and China

23. In their introductions, Ms. Ahn stressed the importance of data warehouses as integration tools, and Mr. Survo thanked Statistics Korea for their financial support for many of the Asian participants.

24. The representative of Cambodia presented their ICT infrastructure strategy in the production of official statistics, dissemination and data warehousing. To meet the increasing demand for good quality statistics the National Institute of Statistics (NIS) undertook a reorganization which included the creation of an ICT department. The new strategy in the production and dissemination of statistics included a standardized network, data storage, limiting the range of software, and developing dissemination tools, including for the on-line publication of census results. The NIS is planning to implement SDMX standards. They use the SuperSTAR software family internally, and plan to use SuperWEB and Redatam On-Line for disseminating census data in the future. Currently they use a customised version of DevInfo (CamInfo) for disseminating MDG and census data. The strategy needs to support the key priorities in a cost-effective way and must have a long-term plan. Staff should have a general knowledge of each area but specialist competence in their own field.

25. The representative of India presented their experience in setting up a national data warehouse. They have a very decentralized statistical system. Official statistics on various subjects are produced by different Ministries and Government agencies both centrally and in the states. The National Statistical Office plays a central and coordinating role among the various agencies. The objectives of the National Data Warehouse are: to develop an integrated repository of current and historical data; to encompass data generated by various Government agencies at the Centre and Provinces; to build a state-of-the-art data warehouse with

online analytical processing (OLAP) capabilities; to enable web-based access to the Data Warehouse. Access is provided on two levels: General users have free and complete access to all metadata, published reports and pre-formatted tables from different surveys and censuses, whilst authorized users are allowed to generate dynamic tables from the data available in the warehouse. Among the difficulties encountered are licensing restrictions in the use of the software, and recruiting and retaining good IT staff.

26. The delegate from Bhutan presented the Internet data collection system of the Labour Market Information Division of the Department of Employment. They have developed a system for sharing and disseminating information in close cooperation with other ministries, departments and agencies involved in the collection and compilation of labour and employment information. The system seeks to generate more reliable, timely and internationally comparable data, establish closer collaboration among policy makers, employers, job seekers and others in the production, use and dissemination of labour market information, and facilitate the formulation of effective labour market policies.

27. Statistics Korea presented the Statistical Self-portrait visualization tool, which addresses the close relationship between statistics, knowledge and policy. The needs of experts and policy makers as well as the general public should be considered when data are disseminated. Statisticians should therefore try to present information meaningfully instead of just presenting survey results. Advanced visualization services should be used when displaying data to help users understand the information presented. This user-oriented service contributed to a substantial improvement in the KOSTAT image. The future direction of statistical information services should be customization and visualization, to transform statistics into knowledge.

28. China presented the application of remote sensing technology to measure crop acreage and yield. The current survey system has problems with quality, and it proved necessary to improve the crop sampling system. Acreage should be reported annually and yield three times. The National Statistical Remote-sensing system of Crop Production (NSRCP) uses a geo-spatial framework, ground-support network, planting area measurement, growth monitoring and yield estimation functions. Testing shows a need to improve the ground support network by setting up a wireless sensor network, and increase the accuracy of crop acreage and yield estimation. Remote sensing data from many sources is used, including from French and Indian satellites.

29. The representative of UNESCAP summarized the common challenges in managing statistical information systems in Asian countries. All countries have made huge progress and IT offers more to statistics than ever before. The Millennium Development Goals have had a very positive influence on statistical systems and cooperation between all parts of national statistical systems. The link between data collection and its primary purpose, better policy making, is increasingly well understood in many countries. There is a wide adoption of certain technologies such as data capture for censuses. Metadata still play a secondary role in most Asian countries, though many statistical organizations in the region have made progress in understanding the central role of metadata in developing their services. Several have started documenting their survey and census data sets in accordance with DDI using the Metadata Management Toolkit of the International Household Survey Network. Systematic data exchange is not only for advanced statistical offices, it is one of the most effective ways of addressing national coordination problems and promoting common methodology in national statistical systems. Recruitment and retention of staff is difficult as public sector salaries are often not competitive.

30. The following points were raised during the discussion:

- Many of the issues raised are universal, not confined to just developed or developing countries.
- Capacity building materials already exist in several places, including the self-learning tutorials being developed for the SDMX website.
- Donor influence can be an issue. Donors often try to push particular software or methods. More coordination between donors would help, and countries should develop their IT strategy before dealing with donors. It is important to decide on standards before starting to build a coherent system.
- IT strategies should follow on from business strategies.

- Sharing experiences is very important to be able to decide on which software to invest in, but the cost of attending meetings is a barrier for many countries. MSIS should continue to investigate supplementary approaches such as the MSIS Wiki, discussion forums, webcasts and recording of meetings.
- If national architectures are studied, it might be possible to derive a small number of standard patterns, taking into account factors such as the size of a country and the degree of centralization in the national statistical system.
- Architectures should be standards-based, this can help influence commercial software suppliers to support these standards.
- Networks of statistical IT staff are important, but need to be facilitated by international organizations.
- The topic of modernizing statistical information systems will be proposed for discussion by the ESCAP Statistical Committee.
- Data management is an important aspect that should be discussed at a future meeting.

Topic (iv): Innovation and related issues

Session Organizers/Discussants: Trevor Fletcher (OECD) and Jolanta Stefanska (IMF)

Papers by: Cambodia, Sri Lanka, Thailand, Indonesia, Canada, Italy/Eurostat, Russian Federation, Australia, United Kingdom, Bank for International Settlements and Eurostat.

31. This session reviewed innovative new developments in statistical information systems covering all aspects of the statistical data life cycle (data capture and validation, data storage, dissemination, data exchange, graphical interfaces etc.) and also developments related to the 2010 round of population and housing censuses.

32. The National Institute of Statistics for Cambodia presented its IT strategy for the 2008 population census. Various constraints during the pre-census and post-census activities were encountered including the lack of trained IT staff, computer viruses, coding errors, inconsistent data and administrative area boundary changes. However, the IT strategy for the census played an important role and served as an essential tool to successfully complete the census work. Manual data entry, using the CSPro software was chosen in preference to scanning and optical character recognition for various reasons including the quality of printing of questionnaires and the familiarity of staff with the manual approach.

33. The representative of Sri Lanka presented their ICT application for the 2011 population and housing census, covering mapping, the enumeration process, computer-assisted manual coding and data processing. The constraints for the mapping phase were the lack of GPS equipment, lack of experienced staff and high cost of the equipment. The enumeration procedures were usually done manually but a pilot e-census project will offer the option of internet returns in selected urban areas. Data capture will be done using image-based form processing technology. There are issues in industrial and occupational coding as descriptions are not always clear and there are problems with translation into native languages. For the next census, on-line editing will be introduced to improve the efficiency and imputation procedures will be explored. Systems are being mainly developed in-house.

34. A multi-modal collection method for the 2010 population and housing census was presented by the representative of Thailand. Paper questionnaires, the Internet and computer-assisted telephone interviews will be used to try to reduce coverage error, and reach hard-to-enumerate households. Scanners will be distributed to regional offices, and data will be sent to the central office electronically via the government network. Dissemination will be via the Internet, including a geographic interface, and micro-data will be made available for researchers via a secure data lab.

35. The representative of Indonesia presented their software systems for surveys and censuses. They have identified which systems should be used, those that already exist, and those that need to be developed. Issues to be addressed include communications, process documentation, document handling, accuracy and the different uses of census data. A general information technology blue-print is needed, and will hopefully

be developed by 2012. Software is usually developed in-house but standards are needed, and will be developed. Outsourcing is not an option for lack of funding.

36. The presentation by the representative of Canada covered innovations in collection management for the 2011 population census. Until 2006 the Canadian census was run as a decentralized operation, due to the geography of the country. In 2006 an internet response option was introduced which led to a complete overhaul of the collection methodology, and a shift to a centralized approach. Time-lags in managing responses, particularly paper questionnaires, resulted in unnecessary follow-up action, increasing respondent burden and causing frustration for field staff. The goal of the new Field Management System (FMS) is to increase the efficiency of collection operations, reduce response burden, decrease clustered non-responses and improve management information. FMS will be integrated into the broader survey collection methodology. The official target is 35% Internet response, though the actual figure may be higher, this will allow a significant reduction in the number of field staff.

37. The European Census Hub Project was presented by Italy on behalf of Eurostat. The aim is to provide users with easy access to detailed census data that are methodologically comparable between European Union member states and structured in the same way. The hub approach offers the most efficient solution to meeting the requirements and the constraints for dissemination at the Eurostat level of 2011 Census data. The SDMX infrastructure built for the Census Hub project can be used in other statistical domains with few or no changes. The most expensive part of the pilot project was the mapping of classifications, but Eurostat has developed a mapping assistant that greatly reduces the time and cost of this work.

38. A delegate from the Russian Federation presented their unified information system for statistical data collection, processing, storage and dissemination. The state statistics service is structured in three levels: federal, regional and district. Rosstat processes and generates 60% of statistical information in the Russian Federation. The main goals are to decrease the data collection and processing period, decrease the labour costs and implement modern information technology. Sub-systems focus on classifications, data collection and processing, and information storage, presentation and dissemination. It takes between two weeks and three months to add new surveys, depending on their complexity, much less than previously.

39. The Australian Bureau of Statistics (ABS) presented their experiences in the use of new technologies such as virtualisation and “cloud” computing and also explored further opportunities for evolving technology infrastructure to support the future of statistical institutions. Drivers for change are similar to those in businesses: statistical organizations have to maintain data quality with shrinking budgets. There was a growing gap between business needs and what IT could deliver. The ABS wanted to move from a hard to manage, low utilization and inflexible “server sprawl” system to a dynamic, computing resource for clients in the form of a virtual cloud. Applications are evenly distributed throughout the servers to prevent slow-downs, and the impact of major technology upgrades is much lower. The biggest applications will be amongst the last to move to the virtual environment, to ensure performance is satisfactory.

40. The presentation by the representative of the United Kingdom Office for National Statistics (ONS) explored the emergence of the Semantic Web, and in particular of Linked Data. The semantic web is a sort of parallel web that sits alongside the “traditional” web. The concept is that everything has its own unique identity in the form of a URI (uniform resource identifier), and can be linked together. The query language used is SPARQL, based on SQL. He examined the relationship between SDMX and Linked Data, and reported on a workshop that the ONS organized in February 2010, which initiated an international collaborative investigation of issues surrounding the publication of statistical data in Linked Data form. This Workshop concluded that there is a need for a Linked Data representation of the SDMX model, and the model is well placed for adoption by those seeking to make data available using Linked Data standards. If we can agree on common datasets we come to a universal naming standard at least for the core things produced. An ultimate consequence of this approach could be the use of RDF (resource description framework) as the format for holding and describing basic data within statistical organizations.

41. Eurostat presented their Statistics Explained project, which is a user-friendly, low-cost dissemination system of statistical information using MediaWiki, an open source software wiki package used by Wikipedia. The system is easy to set up and use and could therefore be attractive for statistical offices of both developing and developed countries. A large amount of value-added content is available in publications but these do not easily reach audiences, so a new approach was needed. Statistics Explained provides a much faster and a more efficient way of disseminating value-added content; is a much better use of the analytical and contextual information available in Eurostat; allows for better integration of all Eurostat output (data, analysis, methodology, cross-cutting topics, etc.); and over time, allows for a substantial quality improvement of all Eurostat value-added output. It uses existing open-source software, which took 15 person-months to install, configure and test. The main version will be in English, but more limited French and German versions are planned.

42. The Bank for International Settlements (BIS) presented their SDMX-based Unified Data Catalogue (UDC) prototype that allows users to search and retrieve data from different types of internal and external data sources. They reported on the technical implementation of the prototype, the conclusions of the feasibility study, and the issues that need to be addressed to move the UDC concept forward. A more powerful registry tool may need to be developed, and the variety of SDMX implementations by different organizations adds complications, though the prototype has demonstrated the feasibility of this approach.

43. The following points were raised in the discussion:

- SDMX self-training modules are available at www.sdmx.org, and more will be added. Eurostat also provide regular training courses for statisticians and IT specialists.
- Many statistical organizations have an innovation function, but this is not always explicitly reflected in the organization structure. Resources are also often diverted to deal with pressing issues with current systems.
- Innovation in IT departments needs to be supported by inputs from the business side and from methodologists.
