

Chairman summary of session IPM 67: Models of modern data and metadata systems

Steven Keuning¹

The design and operation of modern data and metadata systems pose a number of challenges. They require sophisticated information technology (IT) systems, including increasingly complex data warehouses. Systems also need to be able to communicate with one another, which requires the development and implementation of international, preferably global, data exchange standards such as XBRL for accounting information and SDMX for statistical information. The six papers (four on official statistics and two on research) covered the following topics:

- Lessons learnt in developing data and metadata management systems for statistics production and data access and analysis
- Data and metadata management and governance in statistical institutions, from micro to macro and from respondents to final users
- Strategic opportunities from a central metadata store: improved data quality, better management information
- The benefits and opportunities for organisation of data and metadata management within the SDMX service-oriented architecture and data framework
- Web-based technologies for decision-makers and the public in accessing, retrieving and using statistics and metadata
- Making independent data sets relatable through metadata
- Standardising documentation using an international standard in XML and a relational database, for the content and exchange of metadata describing social science data (the Data Documentation Initiative).

Discussant: Lars Thygesen (Statistics Denmark)

The discussant advised participants not to be over-ambitious in setting up metadata systems or statistical information systems; a modular and stepwise approach should be applied, making sure that modules can be used independently. He emphasised the need to see whether solutions developed by other organisations could be used rather than “re-inventing the wheel”. He stressed the need for standardisation within and across organisations and said that it is necessary today to be SDMX-compliant. He argued that developing common data structure definitions (DSDs) was vital to inter-agency standardisation, and that DSDs should be built based on a conceptual model of the reality that statistics need to depict.

¹ European Central Bank.

Concluding remarks

The issues covered today went beyond IT challenges and looked at the impact on data quality, statistical capability, response burden and costs. They can change the way decision-makers and citizens could access and use statistics in the future.

Future work should promote standardisation, encompassing the use of best practices, such as those described in the session, in data and metadata management and governance and in statistical and research institutions, from micro data to macro statistics and from respondents to final users. The vision for databases, input data warehouses and statistical processes lies in an input system in which information is supplied only once and is produced and disseminated in harmonised formats which should be based on widely used international standards (such as SDMX). A vision for IT applications that can read and understand statistical data independently of their location, origin and IT implementation would support re-usability and management and staff expertise and would lower metadata development costs. Web-based technologies will enable these visions for systems using metadata to be realised.

In conclusion, it may be appropriate to organise a similar session in two years' time in order to take stock of any progress made in globally harmonised data structure definitions and to review the possibility and necessity of introducing developments in metadata within official statistics to bioinformatics.