Statistical Journal of the United Nations ECE 10 (1993) 91–95
IOS Press

Introduction – international workshop on statistical metainformation systems

Hans-Joachim LENZ
Freie Universität Berlin, Institut für Statistik und Ökonometrie, Garystr.21, D W-1000 Berlin 33, Germany

Klaus NEUMANN
World Systems Informatics (Consulting and Trading) GmbH, Chausseestrasse 51, D O-1040 Berlin, Germany

Following an initiative originating in the DOSES (Development Of Statistical Expert Systems)-Program of Eurostat, an international workshop was held in Luxembourg from 2–4 February 1993, dedicated exclusively to Statistical Metainformation Systems. It was organised by World Systems(Europe)Ltd. and sponsored by Eurostat. The Scientific Program Committee's goal was to convene experts with different views on metainformation in a variety of statistical fields both in theory and practice.

The workshop can be reviewed in statistical terms as follows:

- About 120 participants from 23 countries attended the workshop.
- The scientific program commenced with a welcome address and an opening lecture, and included eight invited papers, 21 contributed papers, a panel presentation and a general discussion, as well as some technical presentations and demonstrations.
- The program was structured into:
  - the *Opening Session* consisting of the welcome address, the opening lecture and three invited papers on "Metainformation: The State of the Art";
  - *Part A* "Structure and Organization of Metadata" with two invited and seven contributed papers;
  - *Part B* "Design, Development and Implementation of Metainformation Systems" with two invited and eight contributed papers;
  - *Part C* "Requirement Analysis" with six contributed papers.

1. Opening Session

In his *welcome address*, Yves Franchet tackled the manifold facets of the increasing use and importance of metainformation in statistics. He stressed that, without a better understanding and handling of statistical metainformation, neither could the existing and future new technologies be used efficiently nor could the retrieval of statistical information be managed sufficiently, both in terms of time and quality.

In the *opening lecture*, *S. Nordbotten* presented an outline of the historical development and a framework of the needs, types and the gross architecture of metadata systems.

The opening session was completed with invited talks on *Metainformation: The State of the Art*.

M. Podehl recalled the experience of Statistics Canada and outlined some of the user's requirements for statistical metadata. The viewpoint was focused towards the staff of statistical offices and to external clients. The author proposed to further develop metainformation management in a similar way to the progress made with information management.

B. Sundgren presented a conceptual framework for a statistical metainformation system. The semantical aspects were focused in a standardized documentation templet which is flow-oriented in its nature and a metaobject graph similiar to the extended entity-relationship model.

2. Structure and Organization of Metadata

D. Hand investigated the semantical nature of the metadata. The author made a distinction between context-specific and context-free metadata. The latter was studied in depth using several examples from measurement theory.

M.S. Silver demonstrated the importance of footnotes. He proposed to reconsider the role of footnotes and presented an approach for the codification of footnotes.

P. Darius et al. described some results from the DOSES modelling metadata project. The authors analysed the methodology of a system to support statistical processing based on metadata. A conceptual model was presented for formalizing metadata.

3. Design, Development and Implementation

K.A. Froeschl proposed a data model for supporting the semantic integration of data from different sources and defined a table specification language for describing table structures.

E. Malmborg and L. Lisagor reported on the project 'PC-DOC' at Statistics Sweden which supports the documentation of a statistical production system. A gross design was presented and some remarks were made concerning its implementation in a WINDOWS environment.

M. Saijets described the U-File-System which is a tool for the easy storage, retrieval and transfer of micro- and macro-data. It is based on a standardized MVS-file system and a metadata system. Conversion to a PC-file like a Paradox file is possible and conversion to the PC-Axis system is under discussion.

4. Requirement Analysis

G. Deecker et al. concentrated on the issues of how and what metadata should be provided to users of public microdata if it is available on machine readable data files. The authors linked this view to the question of the marketing and advertising of such data.

5. Summary

Not only the welcome address but most of the above mentioned invited and contributed papers confirmed the intentions of the Scientific Program Committee, that is, that there were *very heterogeneous reasons* to focus such a workshop solely on metainformation, e.g.:

1) *Metadata and Metainformation were in the air!*
More and more people were talking about it – recognizing that it was a more or less old topic in a new context, making it necessary to formalize metadata.

2) *Metadata and metainformation were also in the projects!*
For example, most of the projects of the DOSES-Programme confronted the participating research groups with the need not only to describe different types of data but related structures and procedures, etc.

3) *Metadata and metainformation has become a subject for academic research too!*
Studying the problems of data modelling and of the design of statistical information systems required a deeper analysis of the structures, properties and functions of metadata and -information both by the statisticians and the computer scientists.

4) *Metadata and metainformation have proven to be an indispensable prerequisite for the harmonization of statistics.*
Much more so than in the past, the broader scope and the international dimensions of required compatibility, availability and reliability are based on a dynamic documentation of harmonized statistical data.

5) *Metadata and metainformation are a condicio sine qua non for an effective use of modern technologies.*
This is true for the development and application of knowledge-based systems as well as for the use of modern communication technologies (see, for example, UN EDIFACT).

Such a workshop requires some *reflection of the historical background* of the phenomenon on "metainformation". Its history is as old as the history of statistics itself, if we assume metainformation to be any description of content, structure,

relation(s), source(s), quality, etc. This is, of course, no definition but an auxiliary wording for a common understanding. With respect to the term "metainformation", we have to differ between the use of this term in empirical research and in official statistics.

In *empirical research* "metadata" showed up in the statistical packages developed in the mid-sixties to support social research and experimental design. The metadata included lists of variable names, the corresponding set of value labels and the data formats (roughly data types). About the same time international institutions like the World Bank and some other national institutions designed schemes to check the data before inserting it into their file or data-base systems. Consequently, data-entries were introduced assisting the user in editing forms (masks) for entering and checking data using edits (checking rules).

Computer science made the contribution of constraints on data-bases in order to achieve the semantic consistency of a database. The semantic aspects of data were picked up by the A.I. community aiming for semantic reasoning and by the object-oriented programmers. Oldford [1] used the paradigm of object-orientation for the manipulation of metadata and Sato [2] for the description of macro- and metadata.

In *official statistics* the terms "metainformation" and "metadata" were introduced twenty years ago in the context of the development of statistical data bases. Bo Sundgren [3] coined it as part of his infological approach.

At the same time, the international family of EDP-experts in the Statistical Offices had a "golden time" due to the existence of the Computing Research Centre (CRC) of the United Nations in Bratislava. International research teams launched several projects directed towards the design and development of statistical information systems [4]. One of these projects was the basis for the metainformation development as part of the UNDP/ECE-Statistical Computing Project in the eighties, with a METIS-Handbook as an important outcome [5].

This development was supplemented by a lot of corresponding activities in the participating national statistical offices as well as by more or less independent projects exploring the realization of different aspects of metadata handling.

Nowadays, a fruitful cooperation between the METIS-group of the ECE and the DOSES-Program of Eurostat provides a good opportunity to reach a new level, both in coordinating the activities and in transforming the research results into standards and – if possible – into applications.

It was therefore an aim in preparing the workshop, to compose the scientific program in such a way that these different activities could be reflected. This diversity has shown up in the opening session as well as in the overview on the state of the art and in the invited and contributed papers.

The full success of this first international workshop specialized in Statistical Metainformation and presenting so many facets of this subject proves their theoretical and practical importance. A lot of impact was given to ongoing research and a broader application of metainformation in statistics.

References

[1] R.W. Oldford, Object-Oriented Software Representations for Statistical Data, University of Waterloo Technical Report: STAT-87-18, 1987.

[2] H. Sato; A Data Model, Knowledge Base and Natural Language Processing for Sharing a Large Statistical Database, in: Proceedings of the Fourth International Working Conference on Statistical and Scientific Database Management, Roma, Italy, June 1988. Springer Verlag *Lecture Notes in Computer Science*, Vol. 339 (M. Rafanelli, J.C. Klensin, P. Svenson, Editors).

[3] Bo Sundgren, An infological Approach to Data Bases, Urval 7, Statistics Sweden, Stockholm, 1973, pp. 104–105.

[4] Anton Klas, Structuring of the Computerised Information System from the Aspect of Integration, Paper CES/SEM.10/22, Seminar ISIS 1978, Bratislava.

[5] USERS GUIDE to Metainformation Systems in Statistical Offices, ECE/UNDP/SCP/H.4.