

Work Session on Statistical Metadata
(Geneva, Switzerland, 18-20 February 1998)

Item 5 of the provisional agenda

METADATA COEXISTENCE WITH INTRANET AND INTERNET TECHNOLOGIES

Submitted by Infostat, Bratislava, Slovakia ¹

¹ Prepared by V. Kollar and I. Baricic.

I. METADATA AND INTERNET

Users of metadata on Internet

1. We do not distinguish between different categories of users, as has been done in some previous papers on metainformation systems. We consider that people using Internet to search for statistical data are, to some degree, experienced in using computer technology and in statistics. They can use Internet services directly to obtain data. The metainformation system and metadata should increase the possibilities and functionality offered by the existing potential of Internet.

Function of metainformation system in Internet environment

2. In previous years, more functions provided by the metainformation system for the information system itself and its users were defined. Some of these functions can prove useful for Internet technology as well.

Navigation function vs. searching capabilities of Internet services

3. A metainformation system cannot replace the research services offered by Internet. Of course, searching for statistical data using existing search engines such as Altavista or Yahoo can lead to ambiguous results. Some keywords, e.g. 'statistical data', 'GDP' or 'unemployment', can be found on different sites connecting to different areas or fields which are irrelevant to our needs. There is another possibility - to develop and use the 'statistical ring' which connects different sources of statistical data with the help of references to sites of other statistical offices, home pages of international organisations, research institutes and universities, etc.. These references may be present on every site, containing statistical data as **first layer of metadata** on Internet.

4. Statistical offices can provide users with the specific tools for navigation in the form of:

- hierarchical structure of Internet pages;
- text description of statistical information; and
- a specific searching engine for narrower navigation of users.

Integration function

5. This function can be used for processing data from different sources of data. In many cases some data cannot be compared due to methodological inconsistency (different definitions of basic topics) or due to classification inconsistency (use of different classifications, or different levels of the same classification).

Function for interpretation of statistical data

6. The dissemination of statistical data in paper form, or in the form of flat data files on diskettes, cannot be accompanied by the dissemination of data description for economic reasons. It is not possible to produce one statistical table with ten or twenty figures and to produce ten or twenty other pages with a description of these data for economic reasons.

7. In the case of Internet dissemination, it is possible to produce common metadata pages which can be referenced from every page of statistical data. These metadata pages can also be used in some cases as a source of metainformation (classification, definitions, etc.).

II. METADATA AND INTRANET

8. The increasing use of the term 'intranet' forces us to consider the question of whether the function of metadata on Intranet contains some specifics that are distinctive from the function of metadata on Internet. The term Intranet refers to any TCP/IP network that is not connected to the Internet but which uses Internet communication standards and tools to provide information to users on the private network (HTML, hypertext, etc.).

9. The use of Intranet can result in the improvement of communication in a company. Every private company network allows collaboration, communication, team work and shared access to information. But, generally, it is not easy to find the required information in the structure of a common network (LAN or WAN) until we know where the information is located and what it is called. The creation of Intranet servers offers us the possibility of preparing, sorting, publishing, searching for and reading in-house and group-specific information in the same manner as on Internet.

10. With the use of links, users can find required information without knowing the exact position of information on the LAN or WAN network servers. A well-structured Intranet which functions correctly could be of great advantage to every company (office). Each employee has the possibility to find required information quickly and easily. It is possible to disseminate messages to everyone or to only selected co-workers through the defined privileges. Employees have the possibility of selective access to certain types of information.

11. A private company network can be accomplished with Intranet by creating one or more Internet servers (WEB servers). These serve as communication nodes and as a repository of information distributed in company. The number and placement of these servers depends on the intra-office complexity of organisation and geographical distribution of positions of the statistical office. Projecting out of one of these nodes from private offices network to Internet, with the modification of the pages for external users and the definition of access privileges, such an Intranet server can be easily modified as a source of providing information about statistical data, metadata and the statistical office for external users of Internet around the world (official homepage).

12. Separate parts of the statistical office can also communicate through Internet. This is worth doing when most branches of the statistical office are not connected directly with private company network. Implied security aspects should be resolved first. The answer is to use firewall or other appropriate software solutions (e.g. Altavista Tunnel from Digital).

13. Below is a summary of the benefits of Intranet for statistical offices:

- Support of communication within the scope of the statistical office (vertically and horizontally). Extension of the communication forms adding services as FTP (file transfer protocol), WWW (world-wide-web), virtual conferences (audio, video) on intranet (e.g. Microsoft NetMeeting), etc.;
- Developing documentation (e.g. methodical directives on preparing statistical surveys);
- Substitution of paper-based documentation with its electronic form;
- Access to documents and connected metadata for all users;

- Unification of processes and flows of information during the development and distribution of documents;
- Distribution of various types of metainformation for regional statistical offices, support of distributive processing of statistical data.

III. METADATA, INTERNET, INTRANET - THE STATUS OF USING IN SLOVAK REPUBLIC

Specifics in the Slovak Republic

14. The Statistical Office of the Slovak Republic (SO SR) has established a special division for Internet/Intranet technologies. Though there exists a WWW server, on which selected information and statistical data are provided for the public (www.statistics.sk), it was developed only as a marginal product, lacking a clear conception of the preparation and presentation of statistical data on Internet. If we compare it in a detailed manner with paper [2], we will determine zero compatibility with the paper.

15. The subdivision of SO SR mirrors the administrative division of the Slovak Republic (8 regions, 79 districts), with the localisation of central headquarters (3 offices) in Bratislava. Each regional and central office is connected through private network, with the perspective of completing building network connecting the offices at the district level. Communication in SO SR is made possible by sharing computers and by electronic mail (E-mail) at present.

16. The Institute of Informatics and Statistics (Infostat) collaborates closely with SO SR. Infostat cooperates in developing the so-called "Automated Statistical Information System" (ASIS) for the Statistical Office. Metainformation System (METIS) is one of the main parts of ASIS. An electronic information service (ELIS) also exists in Infostat, offering statistical data requested by different users using various media (floppy disks, CD-ROM) or through the BBS.

17. In the current stage of developing ASIS the main goal is to transfer statistical data from xBase formats to the relational database (Oracle) and to develop front-end applications. They will be used for manipulation with statistical data and metadata - in the preparation phases of the statistical surveys, or in the processing of data from respondents (values) phases, checking consistency of data, or in the phases of storing data into the databases. It was not planned for ASIS to use Internet for work with data and metadata, or for their presentation and publication on Internet yet.

Activities of working group on METIS in the field of developing Intranet

18. The working group on METIS in Infostat developed the first Intranet homepages (referred to as METIS Web). We wanted to make it underline the great possibilities of using of Internet technologies when working with statistical data and metadata. We also wanted to lay the foundations of the future Intranet of SO SR.

19. METIS Web has a hierarchical structure, in which the single WWW pages are horizontally and vertically organised in a logical way. Such a structure allows easier orientation of users, easy modification and filling in additional content without the necessity of modifying the already created structure, and, when necessary, it allows the demarcation of single levels by way of defining various access privileges. Navigation between the pages is possible at any moment with the help of:

- links to the pages in other hierarchical levels (with the help of buttons);
- table of contents (TOC);
- full-text local search mechanism;
- push-technology (ActiveChannel from Microsoft, for browser Internet Explorer 4).

20. The main role of METIS Web is:

- to develop a new communication channel between SO SR and Infostat for developing ASIS;
- to offer a contact information (addresses, e-mail, phone), simplification of working with e-mail (an integration into the WWW pages);
- easy response to several categories of problems with forms;
- to create a discussion web (newsgroups) on such themes as ASIS.METIS, SUSR.METHODICS, etc.;
- to distribute software products developed by Infostat for ASIS and their new version or patches (possibility to download software products from WWW pages);
- to create a repository of information about METIS, and publish it on Intranet for every employee of SO SR (e.g. terminology of METIS, list of publications about METIS, links to other sources of information about METIS, etc.);
- to provide access to metadata stored in a relational database for arbitrary employees of SO SR only via the Internet browser (such as Internet Explorer or Netscape Navigator), without the necessity of installing specialised software (i.e. Oracle client) - we developed our own "IntraBrowser" (intranet browser);
- to publish other useful information for employees of SO SR or Infostat (directives, circulars, invitations, release calendar, schedule of lectures etc.) as requested.

21. For METIS Web we used the Microsoft software products, which are standards on SO SR (MS Windows 95, Internet Explorer, Microsoft Exchange). The wide variety of tools provided by that company (MS Windows NT 4.0, MS IIS 3.0, MS Internet Index Server, MS Proxy Server, MS FrontPage 98) allowed us to rapidly develop the first functional Intranet node (server), which is not only compatible with the general considerations mentioned in the paper [2], but also with other recommendations.

22. In order to make it easier for the users to work with statistical data and metadata on Intranet, and to save users time, we have upgraded software products for ASIS in the following way:

- integration of electronic mail into these products (the possibility to send direct mail with data or metadata from these products to a predefined user without leaving the working environment);
- links to the WWW pages which the user can activate directly from the products;

- Help files, which are composed using HTML format meaning that they can also be placed on Intranet.

IV. THE PERSPECTIVES AND GOALS TO THE FUTURE

23. For further development, we expect new ideas from regional statistical offices which will contribute to the expansion of Intranet of SO SR and to the distribution of metainformation. We expect to create new Intranet servers at these offices. Also, future lectures, training sessions and workshops dedicated to this problem will certainly raise the computer consciousness of employees of SO SR. In Infostat we would like to develop an English version of METIS Web for Internet in such a way as to contribute to international co-operation in the field of metadata. As a matter of fact, efforts should be made to strictly adhere to agreed international standards and unification.

REFERENCES

[1] UN/ECE (1995), "Guidelines for the Modelling of Statistical Data and Metadata". Methodological material. United Nations, New York and Geneva, 1995.

[2] UN/ECE (1995), "Standards for Statistical Metadata on Internet". Methodological material.