

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

ENGLISH ONLY

Work Session on Statistical Output for
Dissemination to Information Media
(Ottawa, Canada, 12 - 15 May 1997)

WORKING PAPER No. 9

MINIMUM STANDARDS FOR METADATA ON THE INTERNET

Paper prepared by the UN/ECE Secretariat

MINIMUM STANDARDS FOR METADATA ON THE INTERNET

At its October 1996 meeting, the meeting decided to ask the Work Session on Meta Information Systems (METIS) to prepare a set of guidelines for minimum standards to be met in putting metadata on the Internet. The Bureau asked for a first draft to be prepared for its meeting in February 1997.

As the METIS Work Session was concurrent with the October Bureau, the Bureau's request was transmitted to it immediately. It was agreed that Canada, Germany, Sweden, OECD and World Systems, Ltd. would assist the secretariat in preparing a note for the Bureau. Following discussion and subsequent correspondence the Group agreed the following:

1. The metadata required to accompany statistics on the Internet is essentially the same as the metadata required for statistics in other media. All information considered necessary in paper publications should also be provided on the Internet.
2. It is convenient to categorize the necessary metadata into two broad groups:
 - (a) metadata which assists users in their search for statistics.
 - (b) metadata which assists statistics users in assessing and interpreting the data.

The parallel information in paper publications are the contents page and index in the case of (a) above, and in the case of (b) it is the sources, methods, definitions and classifications used which are usually explained in footnotes, annex notes and sometimes in separate books of notes in the case of some quarterly and annual publications which change little from year to year.

A few years ago, it was difficult to mix texts and figures in information presented on a monitor screen but current software overcomes these old problems. Similarly, the current software makes it relatively easy to search and browse (as was always possible in paper publications) but it is also made easy to go quickly from a page of data to associated metadata which could be quite voluminous (equivalent to the separate books on sources and methods associated with some paper publications).

Metadata which helps users to search for statistics

As a minimum standard, it should be possible to search for statistics on the basis of theme or subject, and it is desirable to be able to search for statistics also by keyword or by variable.

Having located the desired statistics, it should be made easy to return to a level at which it is possible to search for other statistics.

In order to meet these minimum standards, it is necessary to adopt the most appropriate technical solution in each case but the definition of those techniques (such as hypertext linking or Standard generalized mark-up language) SGML is not necessary.

Metadata needed for interpretation

The minimum standards for metadata which enable statistics to be properly interpreted by exactly the same as they would be for statistics published on paper. In other words the variables and

other terms used must be clearly defined, sources should be quoted and any classifications system used should be made clear.

Again, the techniques employed will depend upon the software being used and need not be standardized, although the SGML could form a standard.

Conclusion

The general conclusion is therefore that the metadata, or information, which should accompany statistics on the Internet are exactly the same as should accompany statistics in other media. Fortunately, modern proprietary software allows such metadata to be made available readily and it is simply a matter of choosing the appropriate software in each case.

It is assumed here that the metadata is required for statistics which are published on the Internet for wide use in much the same way that they are published in other media. The requirements are different for other kinds of data which are transmitted or stored for specialists use. For example, where large quantities of micro-data from a particular survey are transmitted between specialists then their metadata needs may be similarly specialized.
