

THE USE OF INTERNET FOR THE DISSEMINATION OF STATISTICAL INFORMATION

Paper submitted by Eurostat¹

Summary:

Mass media and the general public can use Internet to have access to data, news releases and even statistical publications. Metadata is of major importance for managing the growing amount of statistical information coming from different sources.

Users need to be assisted in the search for data, in their interpretation and when processing them. Eurostat recommends metadata for Internet to be written for general users, with links and references to more specific information for specialists.

This paper provides Eurostat's view on a set of guidelines to be taken into account when disseminating statistical information via Internet. Eurostat is currently developing a new dissemination environment (EDEN) which will allow a better management of available metadata to the general public.

1. The use of Internet for disseminating statistical information is more and more common, in Europe and elsewhere. According to UN/ECE, more than thirty Central Statistical Offices in Europe have their own homepage on Internet. This means the general public and mass media can have - via Internet - a direct access to data, news releases and even statistical publications. Some countries and international organisations already allow access to statistical databases and provide search facilities. With new technological standards, currently under development, Internet will be more and more the core of all Statistical Offices' dissemination policies.

Concern about quality and comparability

2. The amount of available information, in a specific site as well as in different locations, increases the need for complementary information - or *metadata* - which helps the user to search, retrieve, understand and use the data properly. This is part of Eurostat's mission statement "to provide the

¹ Paper prepared by Mr. Marco Pellegrino, Eurostat

European Union with a high-quality statistical information system". Quality, in statistics, means reliable data and comparable data and to assess data quality, we need higher metadata coverage.

What does the international community expect from us: harmonization, best practices

3. Eurostat is, at present, a source of harmonized and multilingual metadata for the whole European statistical system (ESS). Of course, harmonization is the result of a cooperation between Eurostat and the other Statistical Offices of the ESS, where Eurostat plays a coordination role. There are good examples: the European System of Accounts, the Monetary Union convergence criteria, the Harmonized Price Index and the Labour Force Survey, but more work has to be done in other fields of European statistics.

4. Another aspect is also important: the adoption of commonly agreed "best practices" could help several national statistical institutes to improve the quality of dissemination policies so as to help, in the end, the general public.

A coordinated action for Metadata on the Internet: UN/ECE Work Sessions.

5. The subject was discussed within the Conference of European Statisticians (CES) during the 1997 plenary meeting in Geneva, where Eurostat was present as an observer. In that occasion, UN/ECE member countries agreed that:

- a) metadata on Internet should be considered primarily in relation to the needs of users, including many non-traditional users of statistical data;
- b) harmonization of concepts and definitions is essential, in order to obtain comparable statistics;
- c) national statistical offices often lack the required capacity and they are not in the position to develop standards in this area.

6. The discussion was brought forward by the UN Work Sessions on Statistical Meta-Information Systems (METIS), where a set of "guidelines for metadata on Internet" was prepared with the assistance of a working group composed of Canada, USA, EFTA, Eurostat, OECD, UNSD and the UN/ECE Secretariat.

7. Such guidelines concern the whole field of statistical dissemination, for all kind of users, not only for mass media. Nevertheless, some of the recommendations and the attention given to user groups may provide the Work Session on Information Media with a positive contribution.

General guidelines for both experts and general public

8. The guidelines are aimed at promoting a consistent interpretation of statistics from different sources, by considering data quality and international comparability as strategic issues. Some good practices need to be developed and supplied to all agencies willing to accept them. The implementation of any guideline depends, ultimately, on our ability to meet the requirements of the users.

9. The organisation of existing databases for the dissemination on Internet should be considered as a quality challenge. The set-up of appropriate tools for the organisation of web pages and the management of hyperlinks can lead to an overall improvement of the whole information base, affecting all the dissemination channels.

10. As World-Wide-Web is a global information framework, guidelines have to be carefully addressed to the international community. The knowledge of national environments can vary significantly within this audience. This means that statistical information supplied on the Internet need to be strongly

supported by metadata. A higher degree of information is generally required to assess data quality and to help international comparability.

11. Metadata for Internet should be written for general consumption, in easily readable pages and avoiding any counterproductive information overload. Hyperlinks should be used, as much as possible, to provide specialised users with more detailed information.

User groups

12. A major effort should be made to understand who the users we address are (general public and mass media, subject-matter experts, statisticians), what they need, why (for which purposes) and, finally, how such requirements can be met by data suppliers.

A difference can exist between user friendliness and user-orientation, between what users want and what they really need. We cannot simply rely on user's opinion: some users are extremely reluctant to accept even the basic information they need to make a correct interpretation. Information suppliers are sometimes in the position of trying to help somebody who does not want to be helped.

13. A minimum set of metadata can be set up with reference to user groups:

- **general audience, including mass media:** users that are either not expert, or subject to strict time constraints for retrieving and using the information;
- **expert users:** skilled in searching, retrieving, assessing quality, interpreting and eventually producing statistical information on their own.

14. All users can either look for a customised service (a sort of pre-defined menu) or ask for the full list of options. As in a good restaurant, prices and waiting times may change, with the second option, but the satisfaction can better be guaranteed. It all depends on quality standards of the restaurant.

Metadata groups

15. Users need to be assisted in their *search* for data. Which data are actually available and how can they be retrieved?

16. Users need to understand the meaning and the limitations to the use of the data. They have to be provided with the key elements for a proper *interpretation* and a quality assessment of the data. Advanced search tools are needed to assist navigation on Internet; the web pages (and the underlying data bases) should be organised accordingly, allowing the user to shift easily between data and information at different levels, for a customised investigation.

17. Expert users and statisticians must be able to assess the quality of statistical data in greater detail, so they can make any further data *processing* reliable and consistent.

Information assisting search and navigation

18. Users search for different purposes and with different levels of understanding. It should be recommended not to limit the search to hierarchical subject-matter classifications, but to include other possibilities, such as free-text search, indexation with synonyms and a thesaurus linking natural terms to statistical tables. As far as subject-matter classifications are concerned, some harmonization would improve navigation facilities.

19. The following elements can be recommended as “best practice” for assisting search:

- table of contents of the WEB site
- frequently Asked Questions (FAQ)
- descriptions of statistical subject areas, with some basic definitions
- overview of the statistical system
- contact persons and e-mail addresses for more information
- reference publications
- release calendars
- feed-back facilities
- software-related metadata (information on the use of any specialised software tool)
- links to other web pages (eg press releases) and to other WWW sites.

20. It is also advisable to provide hyperlinks between statistical tables, general information pages and press releases.

Information assisting interpretation

21. The challenge is to provide users with what they really need to make a correct interpretation, even if users in practice may choose to drop part of the information.

22. The following elements can be recommended, among others, as “best practice” for assisting interpretation:

- content description (statistical population, geographical coverage, observation unit, reference periods, classifications and standards, etc.)
- labels for rows and columns;
- basic definitions (concepts, variables, measurement units,...);
- comparability over time (break in series, missing data,...)
- comparability with alternative sources
- footnotes highlighting specific precautions
- explanation of symbols
- methodological notes and model assumptions
- information on error sources and accuracy
- institutional source for the data
- copyright and restrictions of usage
- links to press releases and general information
- contact points for additional information.

Information assisting processing of data

23. This type of metadata covers the information needed when statistical data are downloaded and transferred for in-depth analysis and further application. Metadata considered as “mandatory” for the interpretation should be always available (at least in form of reference address) when tables are downloaded for further processing. In addition, more detailed information on data quality and possible restrictions for re-use can be recommended.

24. Data and metadata should be adapted to further processing in suitable tools (spreadsheets, databases, software packages for statistical analysis, etc.) provided that some information about the software and any specific format used is available.

Management and design issues

25. Management issues and technical problems associated with the design of web sites are often neglected, in spite of the fact that they can strongly affect the practical use of the site for searching and downloading statistical information. The European Commission - DG X - has already edited a Guide for "Information Providers" with reference to Intranet and Internet servers.

26. A draft list of general issues should include the following subjects:

- web management
- user monitoring and feedback
- multiple language versions
- copyright
- metadata generated automatically (date, time)
- webmaster
- *What's new*
- *FAQ*
- graphics and special icons
- last update
- title and Keywords
- links within documents
- links with related sites

Eurostat on the Internet

27. Eurostat adopted an "electronic first" policy for its own dissemination. The latest projects provide for a more intensive use of Internet, as a privileged channel for the dissemination of statistical data and general information.

28. The latest development will probably be called EDEN (Eurostat Dissemination Environment). We are currently developing a new dissemination framework, aimed at supplying more information and more user-friendliness to statisticians and general public. The new environment will be common to all Eurostat electronic dissemination products and will allow a better management of available metadata. Technical specifications for EDEN are still in progress, so that we cannot go into details. The general philosophy is that the new environment will take into special consideration the distinction between two user profiles: general public and experts. General public will be assisted by a "guided access" and several pre-defined search profiles. The shift from the "general" mode to the "expert" mode will give the chance to access the information base in a deeper way, without losing the user friendliness approach.

29. For further details on Eurostat's new web site, please read Working paper 14, New developments in the use of Internet for dissemination of statistical information, *Eurostat 2.0*, by Annika Östergren (Eurostat).