

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Statistical Metadata

(Geneva, Switzerland, 22-24 September 1999)

Topic (iii): Review of experiences with implementing the “Guidelines for statistical metadata on the Internet”

“GUIDELINES FOR STATISTICAL METADATA ON THE INTERNET”

Report by the National Institute of Statistics, Portugal¹

I. INTRODUCTION

1. The paper contributed by Statistics Norway suggests a short summary of experiences concerning the implementation of the “Guidelines”. In our contribution, we start with a brief overview of that paper and point out every feature (concerning metadata) that our web site has already implemented.

2. During 1997, the National Institute of Statistics of Portugal (INE) reviewed the structure of its web site and organized the delivered information in two areas: promotional and Infoline - an on-line information service. Infoline offers the possibility of consulting on-line, a wide range of statistical information tables, graphics, analysis, and other material available for paper-based publications (thus, metadata are also disseminated over the Internet). Infoline allows access to statistical databases through dynamic tables which are constructed at the request of the user.

3. Almost all of the principles outlined in the “Guidelines” have already been adopted by INE (as shown in the following tables). We present, at the end, some features that we think should be added or changed in the “Guidelines”.

II. GENERAL OVERVIEW

4. As mentioned in para. 2 above, this paper “proposes some guidelines for minimum standards for statistical metadata on the Internet” (...) and “should be regarded more as guidelines than as strict standards”. It is clear that the objectives of the guidelines are:

- to support the www dissemination of statistical data;
- to promote a consistent interpretation of statistics from different sources by considering data quality and international comparability as strategic issues;
- to promote a proper usage and processing of public statistics.

5. It is also clear in this paper that metadata explains the content of statistical data; we won't consider the following issues as metadata:

- metadata for searching at micro data level;
- metadata issues linked to pricing, statistical data confidentiality, etc.

III. THREE DIFFERENT TYPES OF METADATA

6. In para. 18, three types of metadata are presented:

- i) those assisting search and navigation
- ii) those assisting interpretation
- iii) those assisting post-processing of data

¹ Prepared by Ana Lucas.

i) Metadata assisting search and navigation

7. The following table shows those types of metadata considered in i) and the actual status of its implementation in www.ine.pt (Portuguese INE web site):

<i>Types of metadata assisting search and navigation</i>	<i>Present in INE web site</i>
Sitemap/Table of contents	Yes
FAQ	Yes
Site news	Yes
Description of statistical subject areas	Yes
Description of the statistical institution	Yes
Description of the statistical system	Yes
Reference publication	Yes
Contact persons	Yes
Release calendar	Yes
Links to other WWW statistical sites	Yes
Feedback facilities	Yes
Press releases	Yes
A list of key words	No²
Local search engine based on free text search	Yes³
A hierarchical subject matter classification	Yes

ii) Metadata assisting interpretation

8. The following table shows those types of metadata considered in ii) (concerning the presentation of a statistical indicator or a separate figure, graph or table) and the actual status of its implementation in www.ine.pt (Portuguese INE web site):

² Not yet implemented; a list containing the most searched keywords (as well as each related synonymous) is going to be created very soon.

³ Searches table headers only.

<i>Types of metadata assisting interpretation</i>	<i>Present in INE web site</i>
Title/content description, including:	
Statistical population	Yes
Geographical coverage	Yes
Observation unit	Yes
Classifications and standard applied	Yes
Labels for rows/columns in tables and elements of graph	Yes
Definitions of labels	
Measurement unit	Yes
Time reference/period	yes
Regional units	Yes
Comparability over time (break in series)	⁴
Footnotes highlighting specific precautions	yes
Source of the data	Yes
Explanation of standard symbols in tables	Yes
Any information on copyright, restrictions on usage	No
Contact points for additional information	Yes
Comparability with alternative sources	No
Links to press releases/summary of findings	Yes
Description of methods used in collection, revision, calculation, and estimation of the statistics	Yes⁵
Information on error sources and accuracy of the statistics	Yes³
Description of background and purpose of the statistics; concepts, variables and standards used	Yes³

Minimum set of metadata required for the correct interpretation of statistics

Recommended metadata for better accessing the quality and comparability of statistics

As said in § 28, this kind of metadata should either be attached to the downloadable data or should be easy to access and download in a separate operation.

⁴ We have some problems concerning comparability in some projects. However, in our *time series database*, the user can collate the right time series, by reading the proper metadata.

⁵ In metadata pages (available through the search by hierarchical subject matter classification) the user can find some documents that contain information which supported specific statistical projects (methodology, sampling, analysis, etc).

iii) Metadata assisting post-processing

9. This type of metadata should be included when statistical data are downloaded for further application. Some tips for the download formats are suggested in the paper.

<i>Types of format for download post-processing metadata</i>	<i>Present in INE web site</i>
Tables or text other than in picture formats (gif, jpg are not practical for further analysis and processing)	Yes
General and open formats (comma separated, ASCII, etc) should always be present	No⁶
Proprietary formats (spreadsheets, statistical packages) may be useful to simplify further processing and include basic metadata	Yes

iv) Other recommendations

10. This paper also recommends to adopt or develop some guidelines concerning:
- the design of pages, formats for downloading, etc.;
 - management of the web site and allocation of resources;
 - user monitoring: analysis of users' demands;
 - Language: (English version exists?);
 - update information: automated, dynamic links to general databases are often preferable to ad-hoc updating of static pages;
 - What's new;
 - "Title", "Classification" and "Keywords";
 - Country/region name (included in the title of the www home page).

All the items presented above have been adopted by the INE web site. Our English version is a "restricted version" of our web site. It shall be improved very soon.

IV. WHAT SHOULD BE ADDED TO OR CHANGED IN THE "GUIDELINES"?

11. We are currently investing considerable effort in organising metadata concerning methodological issues. One of our goals is to provide detailed methodological metadata to internal and external users with different levels of statistical expertise. We therefore suggest that some highlights could be made of the topic "Recommended Metadata for better assessing the quality and comparability of statistics". We have included in our Intranet system methodological documents related to our statistical operations. These documents include metadata items related to statistical operations such as description, objectives, frequency, sampling methods, sampling estimators, non-responses treatments, statistical concepts and many others.

12. Our documentation is in free text form and it would be useful if there were some guidelines that could provide a common template to organise and disseminate methodological metadata documents. That template should serve two main purposes: to provide a common framework of reference at an international level; and to provide different levels of detail according to the degree of expertise or purpose for which the statistical data is to be used.

⁶ Some tables can be downloaded in CSV (comma separated values) format