

**UNITED NATIONS STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION  
STATISTICAL OFFICE OF THE  
EUROPEAN COMMUNITIES  
(EUROSTAT)**

**ORGANISATION FOR ECONOMIC  
COOPERATION AND DEVELOPMENT  
(OECD)  
STATISTICS DIRECTORATE**

**Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)**  
(Geneva, 9-11 February 2004)

Topic (iii): Metadata models and terminology

## **Integrating Statistical Terminology Tools within Eurostat's Dissemination Policy**

**Contributed Paper**

Submitted by Eurostat <sup>1</sup>

Abstract	<p>During the last METIS session, in March 2002, Eurostat presented a paper on THESEUS, a multilingual thesaurus for accessing Eurostat reference databases. The aim of the present paper is to describe the most recent policy decisions taken by Eurostat with regard to its dissemination policy, and the general impact of these decisions on metadata tools, including THESEUS.</p> <p>The paper emphasises the need for a realistic and international compliant metadata policy. Finally, it gives a short account of some recent developments in metadata management, with special regard to the standardisation of metadata terminology within SDMX (metadata common vocabulary) and to the management of the internal metadata repository.</p>
----------	---

---

<sup>1</sup> Prepared by Bart De Norre, Dominique Groenez and Marco Pellegrino, Eurostat, Directorate for Statistical Methodologies and Tools (<mailto:ESTAT-NewCronos-Metadata@cec.eu.int>).

## I. INTRODUCTION

1. During the last METIS session, in March 2002, Eurostat presented a paper on "THESEUS, a multilingual thesaurus for accessing Eurostat's reference databases"<sup>2</sup>. The aim of the present paper is to describe the revisions taken by Eurostat in its dissemination policy and the general impact of this revision on metadata tools, including THESEUS. The paper will also give account of some recent developments in metadata management, with special regard to the standardisation of metadata terminology and to coordinating the different systems managing textual information.

## II. EUROSTAT'S DISSEMINATION POLICY

2. During the last two quarters of 2003, Eurostat gave full priority to the free dissemination – via Internet – of about 1000 statistical indicators, including *Euroindicators*, *Structural indicators* and a selected set of other long-term indicators. The objective is to extend this number of indicators progressively. For instance, we plan to add as free data Euroindicators' infra-annual time series and breakdowns by country for the following topics: economic statistics on consumer prices, national accounts, balance of payments, external trade, industry, energy, commerce and services, labour market, monetary and financial indicators as well as business and consumer surveys results from the European Commission.

3. Eurostat website now makes this set of indicators freely available in standardised pages<sup>3</sup>, supported by a minimum metadata coverage – in three languages – oriented to the end-user and including title, subtitle, a short description and footnotes. From the standard table view, a hyperlink to "more information" points towards a more comprehensive metadata file, when this is available, according to the SDDS format<sup>4</sup> (the SDDS format has been chosen as the mandatory format for all the additional information available through this hyperlink, for the whole set of indicators).

4. For its future website, Eurostat decided a closer collaboration - within the Commission - with the Office for Official Publications of the European Communities (OPOCE). This Office plays a central role in the dissemination policy of the whole European Commission, providing tools as the global thesaurus EUROVOC and the future EU-Bookshop<sup>5</sup>. EUROVOC is a *generalist* Thesaurus covering all the activities of the European Institutions. It aims at providing search and retrieval functions for all their official publications: this should also hold for Eurostat's publications, especially in the framework of the EU-Bookshop application. The future website is also going to offer more advanced search functions.

## III. RECENT DECISIONS CONCERNING THESEUS

5. The reorganisation of Eurostat resulted in a revision of priorities for the information system development and for metadata management (including the management of the trilingual content of THESEUS). In 2002-2003 Eurostat had finalised the informatics development work on THESEUS: a consultation site on the EUROPA server was ready, together with indexation procedures and search functions exploiting the thesaurus. However, with the reorganisation of Eurostat, contracts and resources were cut and final implementation was postponed. Furthermore, because of the core priorities, existing staff for metadata and information systems needed to focus on the free Internet dissemination.

---

<sup>2</sup> WP 2 (<http://www.unece.org/stats/documents/2002/03/metis/2.e.pdf>) prepared by B. De Norre and D. Groenez

<sup>3</sup> [http://europa.eu.int/comm/eurostat/Public/datashop/print-catalogue/EN?catalogue=Eurostat&service=free\\_downloads](http://europa.eu.int/comm/eurostat/Public/datashop/print-catalogue/EN?catalogue=Eurostat&service=free_downloads)

<sup>4</sup> SDDS = Special Data Dissemination Standard

<sup>5</sup> EU-Bookshop: Internet application for dissemination and selling of official publications

6. For 2004, THESEUS got a "negative priority": there should be no effort for maintaining its content or to test and integrate its associated search functions within Eurostat's website. The THESEUS consultation website on the Commission's EUROPA server should not be put in production. However, at the time of writing, we hope to save the considerable efforts that were put in this tool, for providing the end user with a Thesaurus reflecting more specifically the "statistical world" of Eurostat. During the building phase of THESEUS, special attention had anyway been paid to ensure its compatibility with other similar EU tools, such as EUROVOC and EURODICAUTOM (Terminological Database of the Commission's Translation Service).

7. Eurostat will now consider the following actions:

- to move a selection of the content of THESEUS into EUROVOC;
- to shift the focus of THESEUS towards the "free Internet dissemination" of statistical indicators (whereas, until now, THESEUS was focused on the reference database NEWCRONOS);
- to re-orient and limit the indexation between NEWCRONOS and THESEUS towards the basic index used in the search within the "free Internet dissemination". For 2005, Eurostat will consider whether a specific statistical thesaurus and associated advanced statistical web searching could be beneficial to end users and statisticians, for supporting the dissemination of free data.

#### IV. MANAGING STATISTICAL TERMINOLOGY TOOLS

8. Eurostat is currently putting more emphasis on the coordination of metadata activities in-house and within the European Statistical System. The Board of Directors decided, in 2003, the creation of a Committee for Metadata Coordination (COMET) with the mandate to guide the work, provide directions and to assess achieved results. Within a coordinated metadata framework, special attention is given to the standardisation of metadata terminology and to the coordination and integration in a Reference Metadata Environment of the different systems managing textual information, such as explanatory texts, concepts and definitions (CODED) and data dictionaries<sup>6</sup>.

9. Eurostat is launching a project of "Explanatory Metadata Information System" to address the problems related to the definition, storage and processing of explanatory metadata (such as short definition, data source, coverage,...) within Eurostat. This tool, enabling a centralised management of metadata, will contribute to improving the quality of statistics and, at the same time, it will reduce the actual need of resources. In the long run, this system will support and guide the whole process of explanatory metadata from production towards dissemination. This metadata system takes into account a standard typology of metadata, defined according to needs and priorities of producers and end users, and shall be compliant with the SDDS standard and MCV terminology. For 2004, the purpose is to get a very limited operational tool focused on those statistical indicators which are freely available on the Internet.

10. Our starting point, in line with the result of parallel activities in other statistical institutes and with recent contributions from research projects and from experts<sup>7</sup>, has been that metadata should be captured along the statistical production process, as a side effect of the main process, and that metadata should be re-used rather than invented *from scratch* every time some "explanatory information" is needed for a specific dissemination product.

11. A group of users and producers of metadata systems within Eurostat<sup>8</sup> recently identified a series of main priorities for improving the way metadata are managed. Among these, there are: a) rationalisation of databases (there is currently a diversity of repositories in which different meta-information items are stored); b) stronger connection of data and metadata along the statistical

<sup>6</sup> Data dictionaries should typically be managed by a nomenclature management tool, such as the existing RAMON.

<sup>7</sup> See METIS 2004, WP.3 (Bo Sundgren, *Metadata Systems in Statistical Production Processes*).

<sup>8</sup> The user group met within the CVD initiative (CVD = Cycle de Vie des Données, i.e. Data Life Cycle). See METIS 2004, WP.23 (G. Pongas, F. Vernadat, *Operational Metadata for Federating Statistical reference Systems at Eurostat*) for more details.

production process (metadata should be shared and reused across different datasets and layers); c) standardisation of metadata.

12. Eurostat is particularly involved in the effort – conducted within the SDMX initiative – for the development of a “metadata common vocabulary” (MCV for short) that should provide a common terminology to be used for identifying a reference exchange format of metadata coming from multiple sources and multiple repositories<sup>9</sup>. The MCV is concerned with the elaboration of terminological building blocks – based on metadata standards or best practices – that can be used for describing the collection, processing and dissemination of data, independently from the metadata model adopted. These building blocks, or “atomic items”, are normally used by statistical institutes and metadata experts for deriving a variety of metadata models and specific dissemination outputs. The advantage of using a common terminology lies in the possibility of defining an easier linkage (or mapping) between different models embedded in different information systems. This is also applicable to metadata repositories maintained within the same organisation in multiple formats for historical or administrative reasons<sup>10</sup>. In all those cases, a common *pivot* format for metadata exchange, like the one being developed within the SDMX initiative, would greatly simplify metadata search, re-use, dissemination and “sharing”.

13. IT systems are strategic for handling metadata and for making them available to users and producers. For this reason, the MCV glossary, containing a set of definitions that describes metadata items used by national agencies and international organisations, is already available within the more general Concepts and Definitions Database (CODED)<sup>11</sup> maintained by Eurostat. CODED, available at <http://forum.europa.eu.int/irc/dsis/coded/info/data/coded/en.htm>, contains definitions of key terminology and concepts used within the European Statistical System (ESS). It currently includes more than 4000 concepts and definitions from the following areas of statistics: Economy and Finance; Population and Social Conditions; Industry, Trade and Services; Agriculture and Fisheries; External Trade; Transport; Environment and Energy; Science and Technology. CODED also includes a collection of abbreviations and acronyms commonly used within the European Statistical System, as well as a collection of “statistical terminology”.

14. CODED is an independent information system, built on a database which is designed to store concepts and related texts (for instance legal acts and methodological manuals)<sup>12</sup>. At the same time, CODED is part of the general Eurostat metadata environment. Its first objective is to provide an easily available source of statistical concepts and definitions for all users and producers, but it also acts as a sort of reference point for standardisation and harmonisation. This objective is as well (at least implicitly) included in the concept of the SDMX Metadata Common Vocabulary: by highlighting - and making visible - definitional inconsistencies, the MCV also aims at promoting a consistent use of existing standards.

15. The challenge, for CODED as well as for the MCV, is that the tool is used by metadata users and producers, not only for documentation but also as a production tool, i.e. for re-using the information posted and for integrating the definitions within each organisation’s metadata environment.

---

<sup>9</sup> See METIS 2004, WP.7 (G. Jones, R. Di Calogero, M. Pellegrino, D. Ward, *Inter-agency Cooperation for the dissemination and exchange of standard metadata*).

<sup>10</sup> For instance, Eurostat collects and maintains a series of methodological information items, in domains like national accounts and short-term industrial statistics, on the basis of specific EU regulations.

<sup>11</sup> The OECD, co-producer of the MCV, is doing the same within its general Glossary of Statistical Terms.

<sup>12</sup> For more info on CODED and the database features, the responsible person within Eurostat is Danny Delcambre, [Danny.Delcambre@cec.eu.int](mailto:Danny.Delcambre@cec.eu.int).

## Annexes

## Eurostat new page for access to “Free Downloads” of statistical information



Free Downloads Monday, January 19, 2004

**FREE DOWNLOADS**

Through its website, Eurostat provides access to a range of statistical information you can consult online or download for free. All this information is available only in an electronic format (HTML or PDF).

**Data**

**Long term indicators**

- [Structural indicators](#) from the Commission's Synthesis Report. The indicators cover the domains Employment, Innovation and Research, Economic reform, Social cohesion and Environment.
- [Free data](#). Tables on many areas of life, work, the economy and the environment in the EU can easily be accessed.

**Short term indicators**

- [Free data: Euro-Indicators](#) for the EU-15 and the euro-zone. Approximately 90 economic and social indicators available. Updated daily. Consult the [Release calendar](#).

**Publications available in PDF format**

- Eurostat's expert statisticians analyse their data to make it easier for their users to understand. Everyone can download **all** publications free of charge in PDF format.

**News releases**

News releases contain key figures on the EU-15 and the euro-zone, the Member States and partner countries. Available in PDF format. Consult the [Release calendar](#) for issue dates.

**Statistics in focus**

*Statistics in focus* provides up-to-date summaries of the main results of statistical surveys, studies and analyses. All issues of *Statistics in focus* are available for free only in PDF format.

Issued approximately 200 times per year.

**Eurostat news**

It includes information about the most important events of the European Statistical System, the latest developments of projects.

**Catalogues**

This collection includes:

[http://europa.eu.int/comm/eurostat/Public/datashop/print-catalogue/EN?catalogue=Eurostat&service=free\\_downloads](http://europa.eu.int/comm/eurostat/Public/datashop/print-catalogue/EN?catalogue=Eurostat&service=free_downloads)

Eurostat Home Page: <http://europa.eu.int/comm/eurostat>

## Single entry point to Metadata-related sites, from EUROSTAT home page

The screenshot shows the Eurostat website interface in Microsoft Internet Explorer. The browser's address bar displays the URL: <http://europa.eu.int/comm/eurostat/Public/datashop/print-catalogue/EN?catalogue=Eurostat>. The page header includes the Eurostat logo and navigation links for 'de en fr'. A navigation menu on the left lists 'THEMES', 'PUBLICATIONS', 'DATA', 'METADATA', and 'FREE&ONLINE'. A red arrow points to the 'METADATA' item, which has a dropdown menu open showing 'Classifications and definitions', 'Methods and Explanatory texts', 'Related publications and legal acts', and 'Projects and links'. The main content area features a central banner with the text 'Publications: Free and online'. Below this, there are several content boxes: 'Essential products' with a link to 'Structural indicators: Introduction', 'News Releases' with links to 'Euro-zone external trade surplus 5.0 bn euro', 'Highest proportion of graduates in Science & Engineering in Sweden, Ireland and France', and 'Euro-zone GDP and EU15 GDP up by 0.4%', and 'Eurostat news' with links to 'Sales online not available', 'Short minutes of the working group "Land Use statistics"', and 'Minutes of the Working group "Agricultural accounts and prices"'. The footer of the page displays 'Statistical Office of the European Communities' and the date 'Monday, January 19, 2004'. The Windows taskbar at the bottom shows the 'start' button and several open applications, including 'Inbox - Microsoft Out...', '1 Reminder', 'METIS2004\_WP32-St...', 'CVD Data Life Cycle...', and 'Eurostat:EUROPA - E...'. The system clock shows 'FR' and '15:07'.

<http://europa.eu.int/comm/eurostat/Public/datashop/print-catalogue/EN?catalogue=Eurostat>

## CODED home page

**CODED** The Eurostat Concepts and Definitions Database

Search

de fr

Useful Links

- Eurostat's Classification Server
- Business Methods
- Structural Indicators
- Euro-Indicators
- OECD Glossary of Statistical Terms
- Other Terminology resources on the WWW

Introduction Help

Categories	Number
All	4194
Economy and Finance	1100
Population and Social Conditions	711
Industry, Trade and Services	617
Agriculture and Fisheries	422
External Trade	155
Transport	567
Environment and Energy	158
Science and Technology	186
Abbreviations and Acronyms	292
Statistical Terminology	364

For more information please contact  
[eurostat-methodology@cec.eu.int](mailto:eurostat-methodology@cec.eu.int)

Local intranet

start | Inbox - Microsoft Out... | Microsoft Word | Cybernews - Microsof... | CODED Home - En - ... | 200109Wash | FR | 18:19

<http://forum.europa.eu.int/irc/dsis/coded/info/data/coded/en.htm>