

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

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

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

Work Session on Statistical Metadata (METIS)

(Geneva, Switzerland, 10-12 March 2010)

**THE EUROSTAT METADATA HANDLER:
CRUCIAL FOR HARMONISING STATISTICAL BUSINESS PROCESSES WITHIN THE
EUROPEAN STATISTICAL SYSTEM**

Emmanuel Clement, and August Götzfried, Eurostat

I. INTRODUCTION

1. This document explains the Eurostat Metadata Handler used for the storage and production of different categories of metadata within Eurostat and the European Statistical System.
2. The Eurostat Metadata Handler and its statistical contents are crucial when responding to the new Eurostat vision dealing with the harmonisation and rationalisation of statistical business processes within the European Statistical System¹. For the implementation of this vision the harmonisation of different categories of metadata is a necessary precondition.
3. The Eurostat Metadata Handler is also essential for the further implementation of the SDMX standards and guidelines within the European Statistical System and beyond.

II. WHAT IS THE EUROSTAT METADATA HANDLER?

4. The **Eurostat Metadata Handler** is a Eurostat portal regrouping several IT applications dealing with metadata and making them accessible through one single user interface.
5. The main components of the Eurostat Metadata Handler are:
 - **The Euro SDMX Registry** containing in general SDMX related objects for data and metadata;

¹ Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade (COM 2009 404 final).

- **The National Reference Metadata Editor (NRME)** to be used by countries for the production of different categories of national reference metadata and their transmission to Eurostat; the NRME provides the appropriate SDMX compliant templates to be used.
- **EMIS** used for the production and release of reference metadata at European level at Eurostat;
- **Ramon and Coded** used for the production and dissemination of additional categories of reference metadata within Eurostat (e.g. concepts, definitions, classifications, etc.).

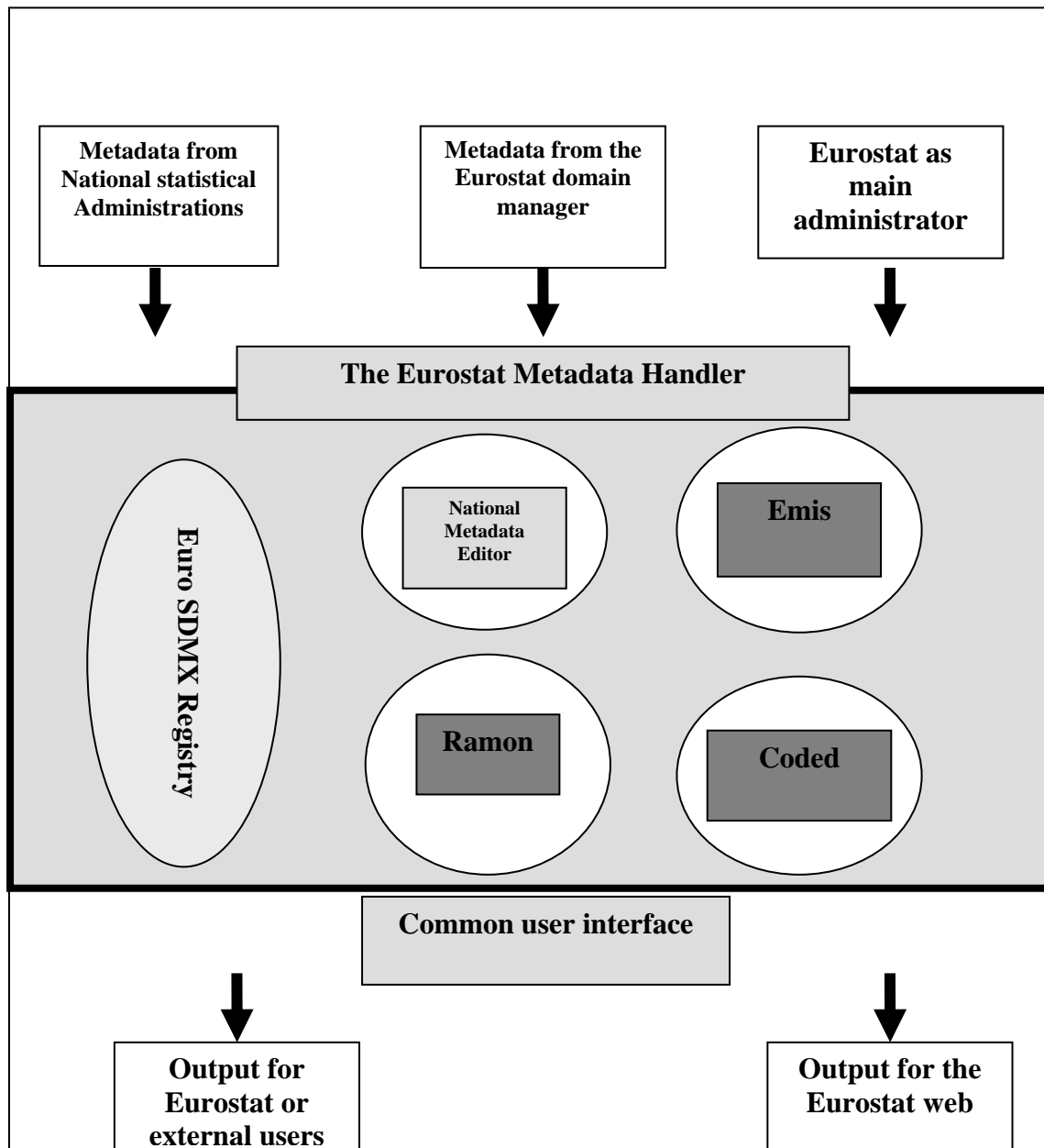


Figure 1: The Eurostat Metadata Handler and its components

6. The main purposes of this Eurostat Metadata Handler are:
- To assure the production and web dissemination of structural and reference metadata within Eurostat and the European Statistical System (other categories of metadata might be added at a later stage);
 - To accommodate and disseminate the national ESS metadata (based on SDMX standards) within the European Statistical System and beyond;

- To stimulate the use of the different categories of metadata for different purposes (e.g. for improving the statistical business processes, for monitoring and improving of the data quality, etc.);
- To provide information on data and metadata flows within the European Statistical System;
- To stimulate the improvement and better integration of the Eurostat and ESS metadata systems in providing more transparency regarding the commonly used and/or harmonised metadata.

7. Over time the Eurostat Metadata Handler should accommodate all categories of metadata which are used from end-to-end of the statistical business process (i.e. from data/metadata collection to data/metadata dissemination).

The Euro SDMX Registry

8. The Euro SDMX Registry is the backbone of the Eurostat Metadata Handler. The registry has the following main contents:

- SDMX compliant data and metadata structure definitions (DSD's and MSD's)
- Harmonised structural metadata (code lists)
- Definition of the concepts used in the different statistical domains
- List of provision agreements.

9. The main purposes of this IT tool are:

- (i) to store information which can be reused by other IT applications being part of the Eurostat Metadata Handler or by other (national) IT applications, and;
- (ii) to make available and disseminate the ESS metadata standards within the ESS and beyond.

10. Most of the contents of the Euro SDMX Registry will be open to users within the European Statistical System and beyond.

The Explanatory Metadata Information System (EMIS)

11. EMIS is the component of the Eurostat Metadata Handler **dealing with the production and dissemination of reference metadata within Eurostat**. This IT component is in use since 2008. Reference metadata are produced internally by the Eurostat domain managers following the Euro SDMX Metadata Structure (ESMS), determined by the Commission Recommendation (2009/498/EC) on reference metadata for the European Statistical System of 23 June 2009. The metadata files produced internally in Eurostat are then disseminated on the Eurostat web site.

12. Additional metadata standards will also be introduced in EMIS (such as the more detailed ESS Standard Quality Report Structure, ESQRS). Moreover also national metadata files (based on standard metadata structures such as the ESMS and the ESQRS) will be inserted into EMIS for further treatment (see also chapter 2.2.3).

13. Technical functionalities have been improved in EMIS in order to enable more flexible extractions of metadata information which can be used for various other purposes (e.g. for working documents, methodological analysis, quality assessments of ESS statistics, etc.).

The National Reference Metadata Editor (NRME)

14. The National Reference Metadata Editor is one of the components of the Eurostat Metadata Handler. It is a Web Application², intended for the production and the transmission of National reference metadata. It allows National Statistical Authorities³ within the ESS to produce national

² An application that is accessed via a web browser over a network such as the Internet or an intranet.

³ Or any other authorised institute, belonging to the ESS.

reference metadata, based on the Euro SDMX Metadata Structure (ESMS) or other reference metadata structures to come, and transmit them to Eurostat via Edamis.

15. Countries can also use the NRME to produce and transmit more detailed quality information, based on the upcoming European Standard Quality Report Structure (ESQRS).

16. At a later stage the NRME will offer its users the possibility to use "tailor" made reporting structures if necessary, e.g. for the reporting of additional categories of metadata.

17. The national reference metadata entered by means of the NRME, are sent to Eurostat via Edamis (the Eurostat Single Entry Point). After processing by Edamis, the data are stored in the Reference Metadata Editor database, from where they can be retrieved by Eurostat Domain Managers, for verification and validation purposes. The latter can then decide whether the Reference Metadata should be published or not on the Eurostat website (pending agreement by the data provider⁴).

18. The National Reference Metadata Editor will be available in the course of 2010.

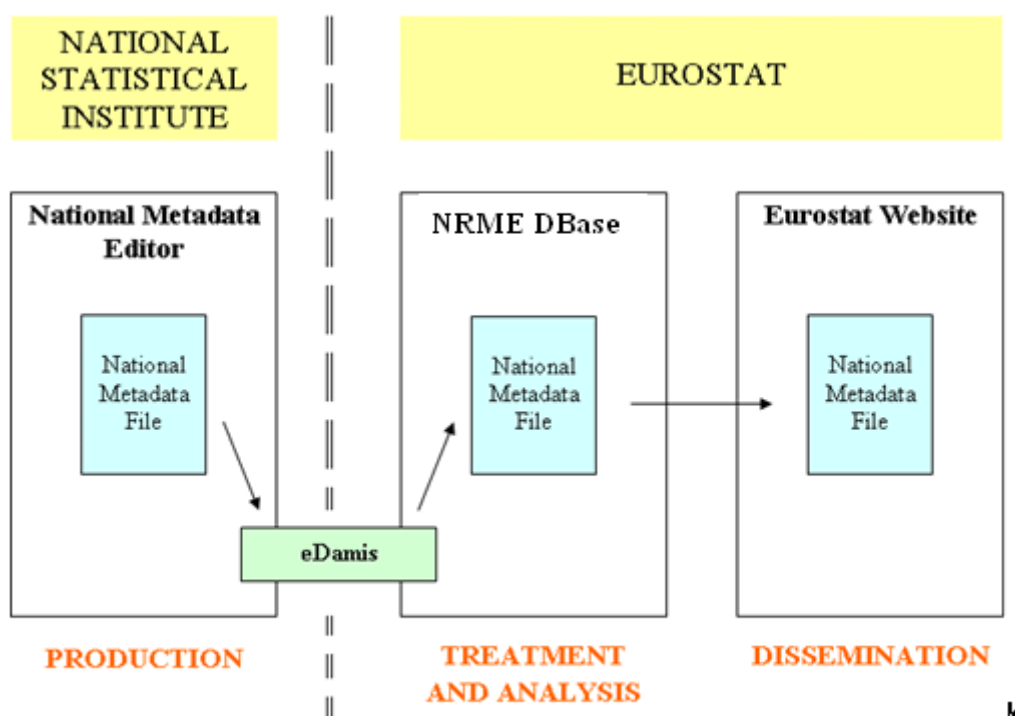


Figure 2: Creation of a National Metadata file

RAMON and CODED

19. RAMON and CODED are the components of the Eurostat Metadata Handler producing and disseminating classifications, statistical concepts, definitions and statistical variables used within the European Statistical System. These two components have a production environment at Eurostat to be used by the Eurostat domains managers. The final output of RAMON and CODED is in general published on the Eurostat website.

20. The description of the various components and functionalities of the Eurostat Metadata Handler is based on the status quo by spring 2010. Further enhancements of the different IT applications and their further integration within the Metadata Handler are necessary. A first candidate for this is the integration of EMIS and the NRME Database.

⁴ Data providers indicate their agreement by setting the "Publication" flag in the NRME to "yes" or "no"

21. In a second step, this might also comprise the further integration of the Eurostat Metadata Handler with the Eurostat IT applications managing publications.

III. ACCESSIBILITY OF THE EUROSTAT METADATA HANDLER

22. The accessibility of the Eurostat Metadata Handler varies according to the different underlying IT applications. The following table gives an overview of the different access rights which will be allocated to the different types of users for each IT component of the Eurostat Metadata Handler:

	Eurostat Domain Managers	National Statistical Authorities	Any public user
SDMX Registry	Access rights for reading and downloading; Central maintenance by ESTAT DB Admin.	Access rights for reading and downloading (with Circa account); uploading of national DSDs, MSDs also possible	Access rights for reading and downloading (with Circa account)
Metadata Editor	Access rights for production and downloading.	Specific production rights for the files related to the NSA and the specific statistical domains (central national administrator)	No access to the application.
EMIS	Access rights for production and downloading; final dissemination of the files centralised.	No access to the application.	No access to the application
CODED/RAMON	Access rights for production and downloading; final dissemination centralised.	No access to the application	No access to the application

23. The Eurostat/ESS structural and reference metadata will be produced via the Eurostat Metadata Handler and will be available on the Eurostat webpage with unrestricted public access, as follows:

- The Eurostat reference metadata files produced in EMIS;
- The national metadata files produced via the National Reference Metadata Editor;
- The harmonised structural metadata (code lists) produced via the SDMX registry;
- The metadata produced via CODED/RAMON.

IV. THE EUROSTAT METADATA HANDLER: CRUCIAL FOR HARMONISATION AND INTEGRATION OF STATISTICAL BUSINESS PROCESSES WITHIN THE EUROPEAN STATISTICAL SYSTEM

24. The Eurostat Metadata Handler and its contents will be essential when it comes to the harmonisation of the statistical business processes within the European Statistical System. As emphasised in the new Eurostat vision, the future statistical and also the IT harmonisation and integration within the European Statistical System will have two directions:

- horizontal i.e. between broader statistical domains and
- vertical between Eurostat and the National Statistical Administrations.

25. The following graph shows this envisaged integration of the European Statistical System:

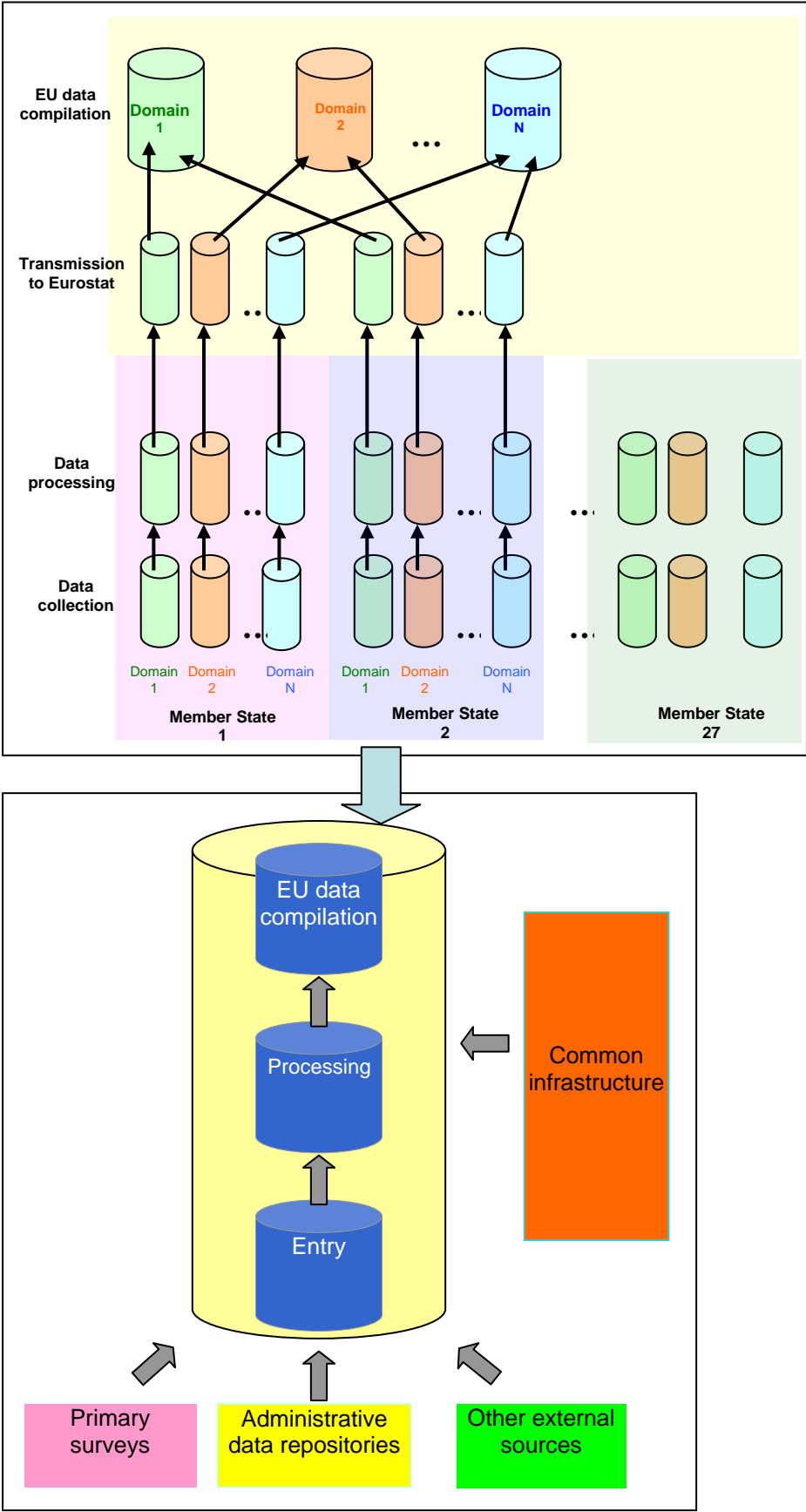


Figure 3: From stove-pipe to more integrated statistical business processes within the European Statistical System

26. The **Generic Statistical Business Process Model (GSBPM)** provides the basic orientation for this further integration of the statistical business processes within the European Statistical System. Whilst National Statistical Administrations are normally dealing with all process steps of this model, Eurostat – as in general receiving aggregated data from national administrations – mostly deals with the process steps 4 to 7 of this model.

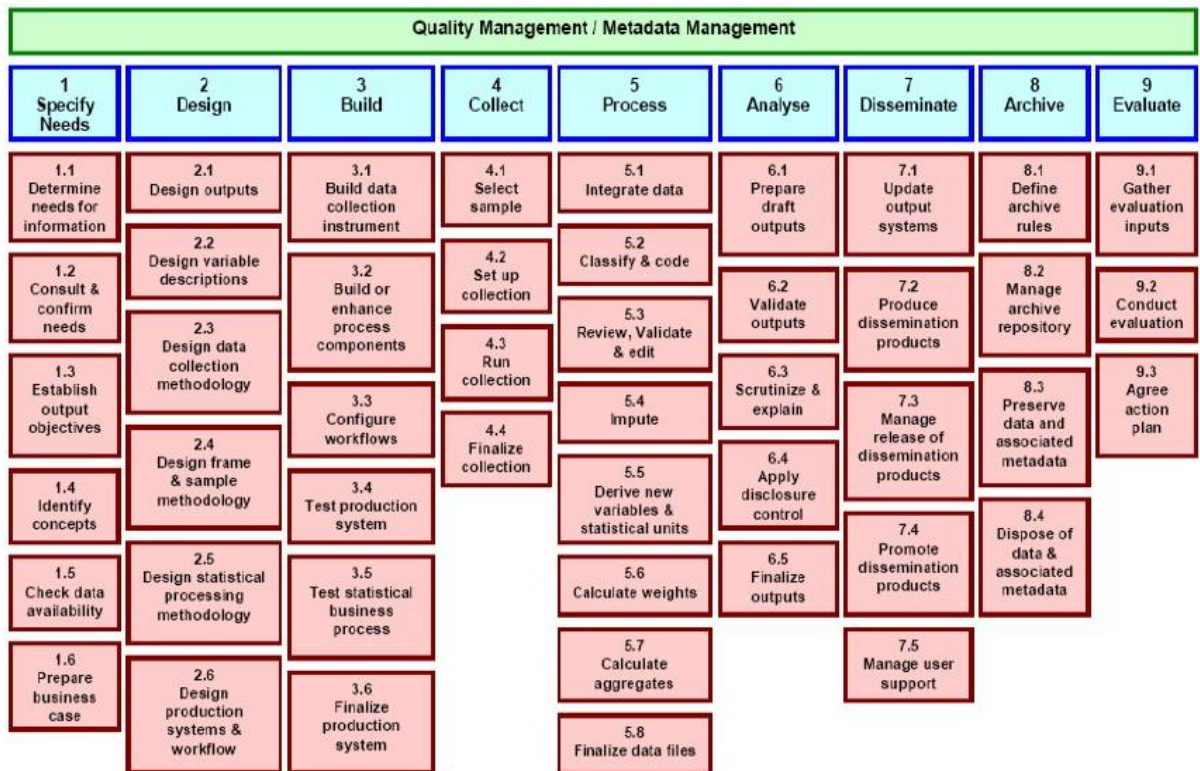


Figure 4: The UNECE/Eurostat/OECD Generic Statistical Business Process Model (GSBPM)

27. The model emphasises the overarching presence of the **metadata and quality management** across the whole statistical data life cycle implying an impact of metadata and metadata management on the different steps of the statistical business process.

28. The GSBPM finds its correspondence in the Eurostat CVD IT architecture (= Eurostat CVD) which is also built all along the data life cycle of the data received, treated and disseminated by Eurostat.

29. As shown in the following Figure 5, the Eurostat Metadata Handler interacts with all steps of the data treatment within the Eurostat CVD. This means more concretely: an interaction with the Single Entry Point (central data/metadata reception point at Eurostat), the Eurostat production databases used in the different statistical domains and with the Eurostat reference database Eurobase.

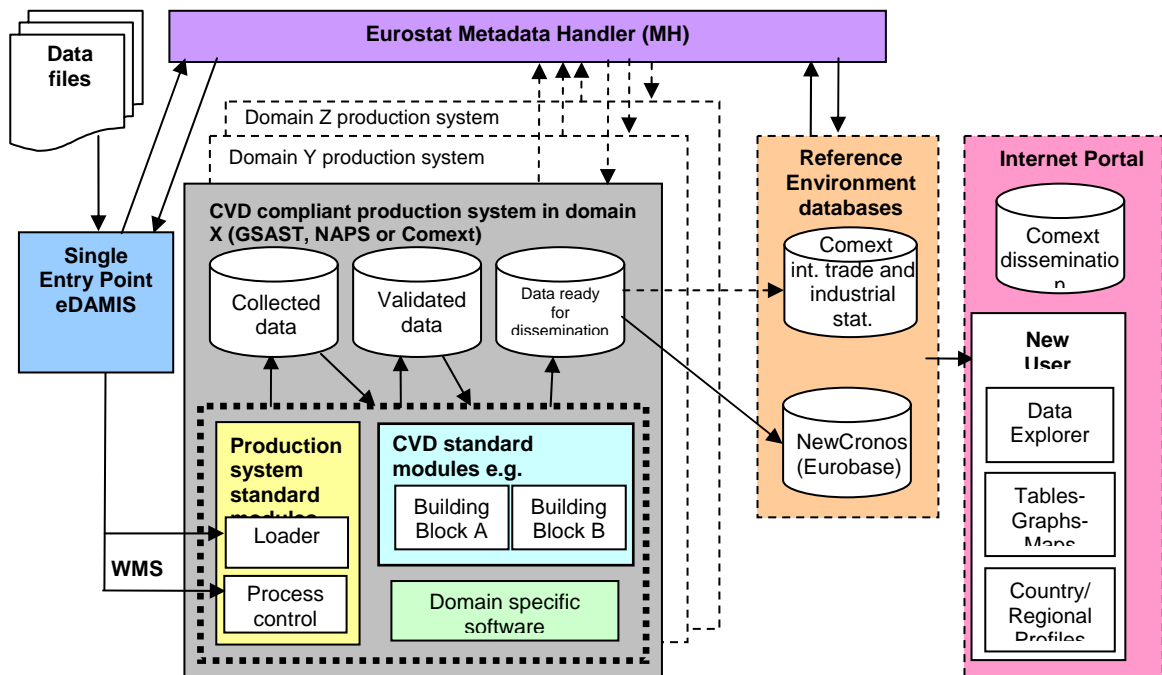


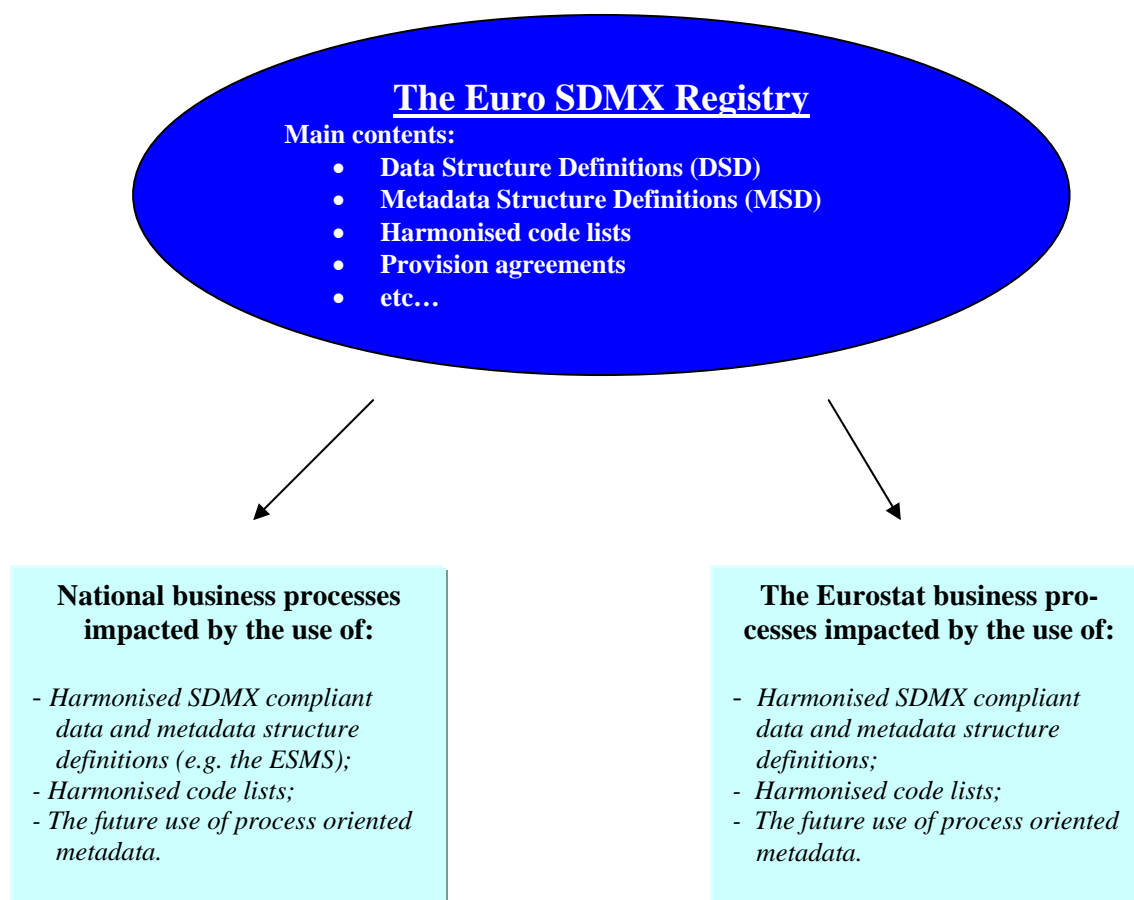
Figure 5: The Eurostat CVD IT architecture (linked to the data life cycle at Eurostat)

30. In the following we concentrate on the two main components of the Eurostat Metadata Handler which impact the statistical business processes at Eurostat and also in the ESS Member States:

The Euro SDMX Registry and the Reference Metadata Editor

31. The statistical and IT standards communicated and used in the Eurostat Metadata Handler are in general fully SDMX compliant. This means that Eurostat fully uses the SDMX data and metadata standards and guidelines and implements them for the European Statistical System when European data and metadata standards are drawn up.

32. The Euro SDMX Registry impacts the statistical business processes at Eurostat and in Member States as follows:



33. The Euro SDMX Registry stimulates the harmonisation and integration of the Eurostat and national statistical business processes as follows:

- Making harmonised code lists available to national statistical administrations and Eurostat: these harmonised code lists should be successively used in more integrated business processes at national and Eurostat levels during all appropriate process steps (i.e. from end-to-end) as outlined in the GSBPM;
- Making SDMX compliant Data and Metadata Structure Definitions available to National Statistical Administrations and Eurostat.

An example: The use of the standard code lists for Economic Activities

Eurostat released the code list for Economic Activities Nace Rev.2 which should be used in data and metadata messages. There the code B_E is used for the activity of "Industry (except construction)". This code should be used in all SDMX compliant data messages using these economic activities when such data messages are compiled within the ESS. The common use of this code stimulates the better integration of the underlying statistical business processes as no additional (unnecessary) derivations are defined and as less and less trans-coding will be necessary.

34. Secondly, the National Metadata Editor also will impact the Eurostat and the national statistical business processes. The National Metadata Editor is a Web application, containing a number of ESS metadata standards, which can be used by National Statistical Administrations for the production and transmission of national metadata.

The National Metadata Editor

Main contents and functionalities:

- The standard metadata structures (ESMS, etc.)
- Use of the application for National Metadata production and transmission
- Existing metadata files (in XML format) can be imported

National business processes impacted by the use of:

- the ESS standard for reference metadata (ESMS)
- the ESS standard for quality related metadata (ESQRS)
- other upcoming European metadata standards

The Eurostat business processes impacted by the use of:

- the ESS standard for reference metadata (ESMS)
- the ESS standard for quality related metadata (ESQR)
- other upcoming European metadata standards

35. **The National Reference Metadata Editor** stimulates the harmonisation and integration of the Eurostat and national statistical business processes as follows:

- In many statistical administrations, the statistical information used for the production of the domain specific metadata using the ESS metadata standards ESMS or the ESQRS now needs to be assembled and comprehensively compiled. Until now this information is often stored in a dispersed and non-harmonised manner within statistical administrations, also using different production processes and file formats. This scattered information will now be integrated and produced via the National Reference Metadata Editor by using the ESMS and ESQR structures. Beyond this, statistical organisations will often re-design their national production processes and databases for arriving at a more efficient production of this information at national level.
- The statistical business processes for handling the incoming ESMS and ESQRS files from countries at Eurostat were rather heterogeneous until now. This not only concerns the heterogeneous contents of the metadata files; it also concerns different file formats used and often unsophisticated business processes for handling these files by the Eurostat domain managers. The new National Reference Metadata Editor will harmonise the metadata file structures, the metadata file contents as well as the statistical business processes used in the Eurostat production units dealing with these national metadata files.

An example: The production of a national metadata file using the Euro SDMX Metadata Structure (ESMS) for Structural Business Statistics in country X

The national domain manager for Structural Business Statistics in country X accesses the National Metadata Editor. In this IT application he/she will be allowed to create metadata files only for his country and only for the Structural Business Statistics (= the domain for which he/she has been defined as being responsible).

The domain manager selects the data/metadata flow for which he wants to create a file. Then, the respective metadata structure (ESMS, ESQRS ...) will be offered to him/her. He/she selects the reference period for his file. Then he/she can fill in the respective metadata template which he has chosen. Once the template is completed, the domain managers can preview the file in html before sending it to Eurostat. This national metadata file will then be transmitted via Edamis to the NRME database and be made accessible to the Eurostat domain manager for Structural Business Statistics. In agreement with the national domain manager, the Eurostat domain manager will be able to publish this file on the Eurostat website.

V. CONCLUSIONS

36. Much has already been achieved with the creation and implementation of the Eurostat Metadata Handler which is successively put in place and used within the European Statistical System. Main next steps of work are the further creation of metadata standards within the ESS (based on the SDMX standards and guidelines) as well as the implementation of the ESS metadata standards in the Eurostat Metadata Handler.

37. Good work progress on these issues is however essential in order to further harmonise and integrate the statistical business processes within the European Statistical System.