

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN
STATISTICIANS

WP. 19
ENGLISH ONLY
EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD) STATISTICS DIRECTORATE

Workshop on Statistical Metadata (METIS)

(Geneva, Switzerland 5-7 October 2011)

Session 5: Topic I: Metadata flows within the GSBPM

Standards and processes for integrating metadata in the European Statistical System

A. Götzfried, H. Linden, and E. Clement

Eurostat

Joseph Bech building,

5 Rue Alphonse Weicker,

L-2721 Luxembourg

August.Goetzfried@ec.europa.eu

KEY WORDS: Euro-SDMX Metadata Structure (ESMS), National Reference Metadata Editor (NRME), SDMX and cross-cutting legislation.

I. Introduction

1. In its Commission Communication "Re-engineering the Production Systems of European Statistics: a Vision for the next decade"¹, Eurostat presented a vision for improving the production methods of European statistics and their efficiency within the European Statistical System (ESS). In May 2010, the European Statistical System Committee (ESSC) presented a joint ESS strategy² for the implementation of this Commission Communication, providing the strategic orientations for the work of the ESS and how it would be possible to modernise this system and equip it in the medium and long term to meet the challenges that statistical producers are faced with at national as well as at European level.

2. As one response to this Vision, the availability of high quality reference metadata for European statistics is considered as an essential pre-condition for the further integration of business processes with the ESS. In accordance to the Generic Statistical Business Process Model (GSBPM), metadata are used as an over-arching bridge all across the statistical business process. Therefore, the harmonisation of metadata within the European Statistical System becomes crucial when it comes to the implementation of the abovementioned ESS Vision. Metadata harmonisation will not only rationalise the statistical business processes used, but also enhance the quality of the data and metadata produced and disseminated within the ESS.

¹ COM (2009) 404, 8 August 2009

² ESSC 2010/05/6/EN, 20 May 2010

3. In order to support the further harmonisation and standardisation, Eurostat is developing SDMX based metadata standards. The Euro-SDMX Metadata Structure (ESMS) is already broadly used at the European level for the collection and dissemination of reference metadata in general. The ESS Standard for Quality Reports Structure (ESQRS) has been developed recently for supporting the more comprehensive quality reporting.

4. Eurostat now aims at using these standards for the collection of national reference metadata in general and metadata targeting data quality. The implementation of ESMS is supported by a Commission Recommendation³ and the implementation of ESQRS through domain related EP/Council Regulations and Commission Regulations requesting the evaluation of data quality.

5. The process of collection of national metadata and their conversion/alignment to the existing standards has already started.

6. In addition to the standards per se, Eurostat also recently made available an IT application for the management of the metadata information from their creation until their dissemination (the National Reference Metadata Editor). Indeed, this SDMX-compliant IT tool will not only enable National Statistical Institutes (NSIs) to produce metadata files but also to transmit them directly through this application to Eurostat. Eurostat, with the approval of the NSIs, will then also be able to broadly disseminate the metadata on its website together with the related statistical data.

7. Finally, and still within the objective of stimulating the further integration of statistical business processes, Eurostat is elaborating cross-cutting and broader framework legislations in order to progressively replace the existing legal acts merely designed as stove-pipes. In this framework, two EP/Council Regulations on metadata harmonisation and on data/ metadata exchange have been drafted by Eurostat.

II. The standardisation of reference metadata at European level

8. Together with the provision of European statistical data, Eurostat disseminates reference metadata files. These files, managed by the Eurostat domain managers in the statistical production units, aim at describing the statistical data and provide additional information which can be conceptual, methodological or related to the quality of the disseminated data.

II.a A standard for producing reference metadata

9. Formerly released on the basis of the Special Data Dissemination Standard (SDDS), Eurostat introduced in 2008 a new standard for the production and dissemination of its reference metadata information: the Euro-SDMX Metadata Standard (ESMS).

10. The ESMS is made of 21 concepts and sub concepts (see figure 1) selected from the list of SDMX cross-domain concepts as described in the Annex 1 of the SDMX Content-oriented Guidelines⁴. Thus, the ESMS has the additional advantage to be an SDMX-compliant structure.

³ EC (2009) 498, 23 June 2009

⁴ http://sdmx.org/wp-content/uploads/2009/01/01_sdmx_cog_annex_1_cdc_2009.pdf

Figure 1. The 21 ESMS concepts and their sub-concepts

Concept Name		Concept Name		Concept Name	
1	Contact	7	Confidentiality	15	Timeliness and punctuality
1.1	Contact organisation	7.1	Confidentiality - policy	15.1	Timeliness
1.2	Contact organisation unit	7.2	Confidentiality - data treatment	15.2	Punctuality
1.3	Contact name	8	Release policy	16	Comparability
1.4	Contact person function	8.1	Release calendar	16.1	Comparability - geographical
1.5	Contact mail address	8.2	Release calendar access	16.2	Comparability - over time
1.6	Contact email address	8.3	User access	17	Coherence
1.7	Contact phone number	9	Frequency of dissemination	17.1	Coherence - cross domain
1.8	Contact fax number	10	Dissemination format	17.2	Coherence - internal
2	Metadata update	10.1	News release	18	Cost and burden
2.1	Metadata last certified	10.2	Publications	19	Data revision
2.2	Metadata last posted	10.3	On-line database	19.1	Data revision - policy
2.3	Metadata last update	10.4	Micro-data access	19.2	Data revision - practice
3	Statistical presentation	10.5	Other	20	Statistical processing
3.1	Data description	11	Accessibility of documentation	20.1	Source data
3.2	Classification system	11.1	Documentation on methodology	20.2	Frequency of data collection
3.3	Sector coverage	11.2	Quality documentation	20.3	Data collection
3.4	Statistical concepts and definitions	12	Quality management	20.4	Data validation
3.5	Statistical unit	12.1	Quality assurance	20.5	Data compilation
3.6	Statistical population	12.2	Quality assessment	20.6	Adjustment
3.7	Reference area	13	Relevance	21	Comment
3.8	Time coverage	13.1	User needs		
3.9	Base period	13.2	User satisfaction		
4	Unit of measure	13.3	Completeness		
5	Reference period	14	Accuracy and reliability		
6	Institutional mandate	14.1	Overall accuracy		
6.1	Legal acts and other agreements	14.2	Sampling error		
6.2	Data sharing	14.3	Non-sampling error		
Concept Name		Concept Name		Concept Name	
1	Contact	7	Confidentiality	15	Timeliness and punctuality
1.1	Contact organisation	7.1	Confidentiality - policy	15.1	Timeliness
1.2	Contact organisation unit	7.2	Confidentiality - data treatment	15.2	Punctuality
1.3	Contact name	8	Release policy	16	Comparability
1.4	Contact person function	8.1	Release calendar	16.1	Comparability - geographical
1.5	Contact mail address	8.2	Release calendar access	16.2	Comparability - over time
1.6	Contact email address	8.3	User access	17	Coherence
1.7	Contact phone number	9	Frequency of dissemination	17.1	Coherence - cross domain
1.8	Contact fax number	10	Dissemination format	17.2	Coherence - internal
2	Metadata update	10.1	News release	18	Cost and burden
2.1	Metadata last certified	10.2	Publications	19	Data revision
2.2	Metadata last posted	10.3	On-line database	19.1	Data revision - policy
2.3	Metadata last update	10.4	Micro-data access	19.2	Data revision - practice
3	Statistical presentation	10.5	Other	20	Statistical processing
3.1	Data description	11	Accessibility of documentation	20.1	Source data
3.2	Classification system	11.1	Documentation on methodology	20.2	Frequency of data collection
3.3	Sector coverage	11.2	Quality documentation	20.3	Data collection
3.4	Statistical concepts and definitions	12	Quality management	20.4	Data validation
3.5	Statistical unit	12.1	Quality assurance	20.5	Data compilation
3.6	Statistical population	12.2	Quality assessment	20.6	Adjustment
3.7	Reference area	13	Relevance	21	Comment
3.8	Time coverage	13.1	User needs		
3.9	Base period	13.2	User satisfaction		
4	Unit of measure	13.3	Completeness		
5	Reference period	14	Accuracy and reliability		
6	Institutional mandate	14.1	Overall accuracy		
6.1	Legal acts and other agreements	14.2	Sampling error		
6.2	Data sharing	14.3	Non-sampling error		

11. After a two-year process, all former SDDS files have been converted into this unique standard structure.

12. In addition to the ESMS, Eurostat has also recently developed a report structure which will be used for the production of the more detailed producer-oriented quality reports within the ESS: the ESS Standard Quality Reports Structure (ESQRS). This additional report structure is successively used in statistical domains.

II.b A dedicated IT application used in Eurostat: EMIS

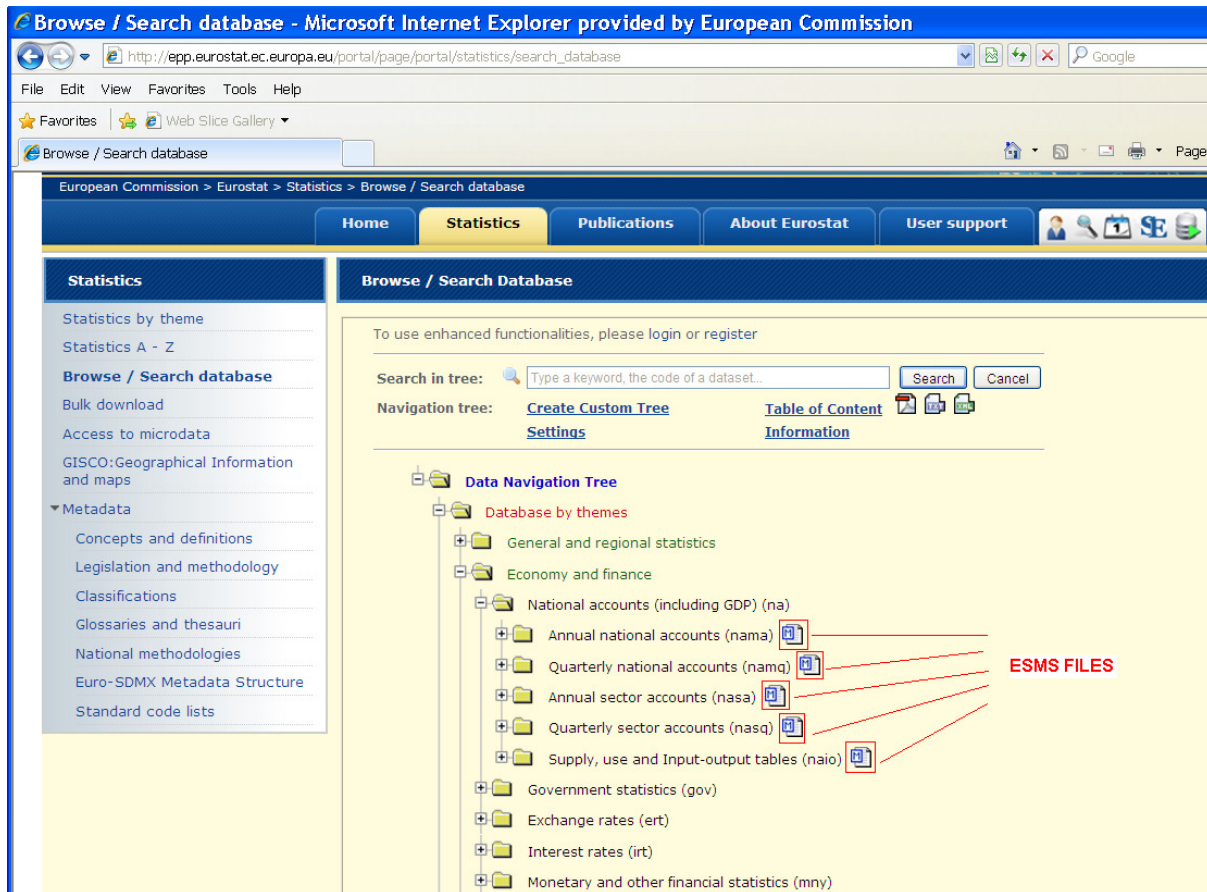
13. In order to create, manage and disseminate the ESMS files, Eurostat developed an IT application especially dedicated to these purposes: the Explanatory Metadata Information System (EMIS). This application is used internally by the Eurostat domain managers who are producing their ESMS files.

14. In order to support the production of these metadata files, Eurostat also developed ESS guidelines which are describing the ESMS structure and explaining how and which type of information to report for each of the 21 concepts and sub-concepts.

15. Once created by the responsible domain managers in the statistical production units, the metadata files are then validated by the Eurostat central unit responsible for metadata who checks the files e.g. in terms of consistency and completeness.

16. The validated files are then disseminated on the Eurostat website together with their corresponding statistical data (see figure 2).

Figure 2. The dissemination of the ESMS files on the Eurostat website



III. Implementation of the ESS standards at national level

17. Since the harmonisation of reference metadata at European level being successful, Eurostat is currently extending this process to the production and exchange of national reference metadata.

18. This has many advantages. First of all, the heterogeneity of the different metadata structures and file formats usually transmitted by NSIs to Eurostat made their production and dissemination cumbersome. The possibility of comparison of information collected from different NSIs was also often reduced.

19. Secondly, the use of harmonised reference metadata at national level perfectly responds to the vertical integration of the statistical business processes within the ESS. One single business process will then be used for national reference metadata from their production at national level to their dissemination at European level, covering mainly phases 4 to 7 of the GSBPM (Collect, Process, Analyse, and Dissemination).

20. Finally, the standardised management of national reference metadata will definitely reduce costs and burden at both NSIs' and Eurostat levels.

III a. Application of the ESS metadata standards

21. In order to be consistent with the reference metadata at European level, the ESMS is also used as basic metadata structure for the production and dissemination of national metadata for all statistical domains. This strengthens the horizontal integration of the ESS business processes further.

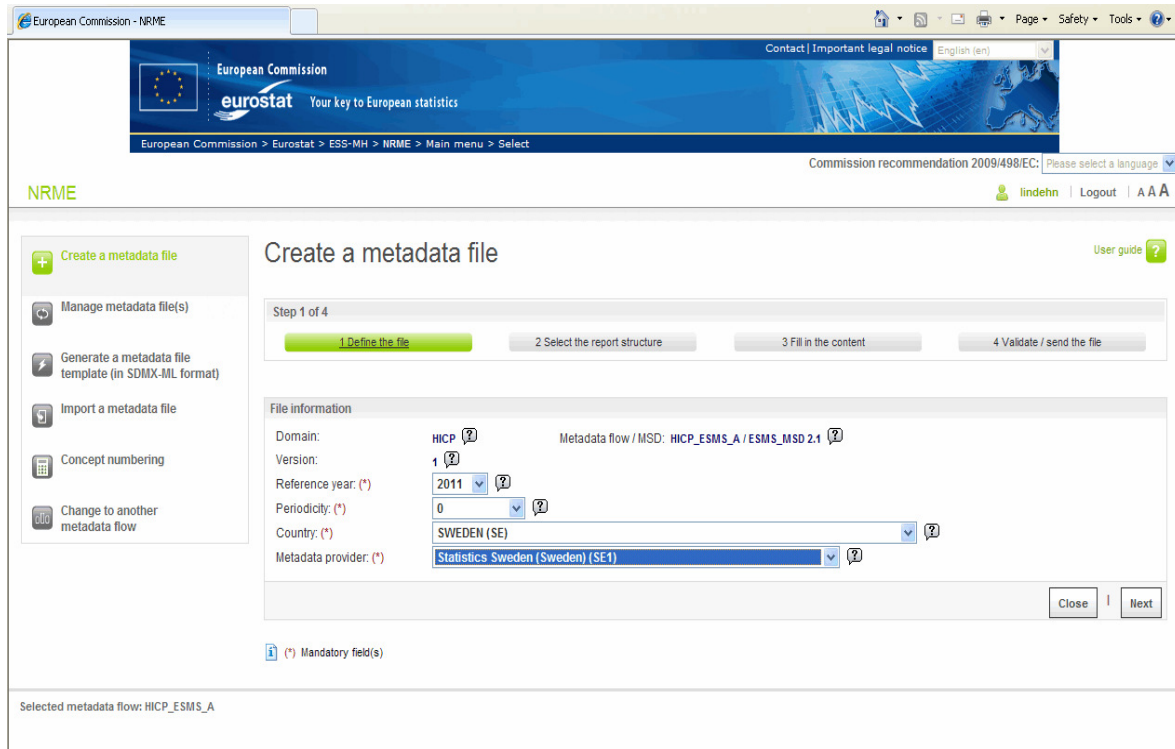
22. The use of the ESMS is also reinforced by the Commission Recommendation 2009/498/EC. Indeed, this legal act recommends NSIs "to apply the statistical concepts and sub concepts listed [...] when reference metadata are compiled in the different statistical areas and when reference metadata are exchanged within the European Statistical System or beyond". The list of the 21 ESMS concepts and their sub-concepts is annexed to the Recommendation. In the meantime, more and more NSIs are also using the ESMS for their own metadata production and dissemination.

23. Eurostat also promotes the further use of the ESQRS for the production and dissemination of harmonised national quality reports within the ESS in line with the recommendations of the ESS Sponsorship on Quality.

III b. The use of a dedicated IT application: the National Reference Metadata Editor

24. To support the production and dissemination of national reference metadata, Eurostat has recently released an IT tool dedicated to the production, exchange and dissemination of national reference metadata: the National Reference Metadata Editor (NRME) – see figure 3.

Figure 3. Overview of the National Reference Metadata Editor (NRME)

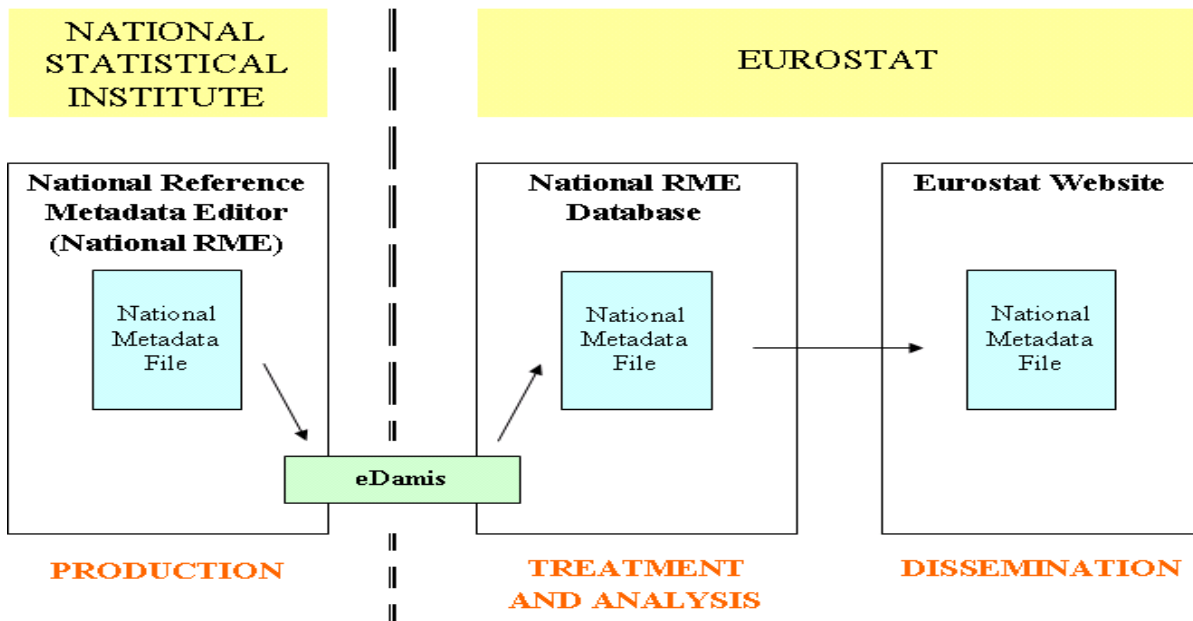


25. The NRME is an SDMX compliant web application offered to NSIs for on-line production and exchange of metadata. Although its use is not a "must" for NSIs, this application has the advantage to design the national metadata process from production to exchange.

26. Indeed, NSIs can not only use the application to produce their national metadata but the information can then also be sent automatically to Eurostat via the single entry point: eDamis.

27. Once stored in an internal Eurostat database, the Eurostat domain managers will then be able to check and validate the transmitted metadata files. In agreement with the national domain managers, the produced ESMS files can then, finally, be disseminated on the Eurostat website (see figure 4).

Figure 4. The process of creation of a national metadata file



28. The NRME allows not only the production and dissemination of national ESMS files. Also national ESQRS files using the same business process can be handled by the NRME.

III c. A first example of a statistical domain implementing the ESMS at national level: HICP

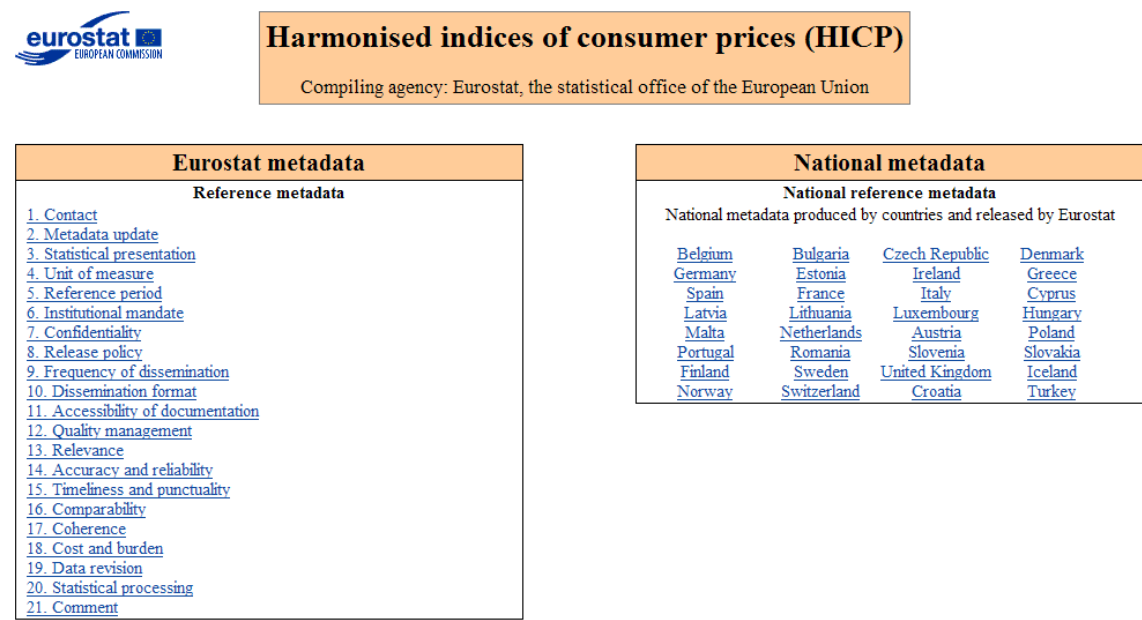
29. In June 2010, Eurostat started to convert the national metadata files in a first statistical domain: the Harmonised index of Consumer Prices (HICP). The HICP used to collect national metadata files according to the SDDS structure and thus the same process as for the conversion of the European SDDS files into ESMS was used for converting the national files into the ESMS structure.

30. After discussion and agreement at the HICP Working Group meeting in October 2010, the conversion process of the existing SDDS files into the ESMS standard started.

31. In November 2010 each NSI received its converted national ESMS file for update and completion. These files were revised at national level and the updated versions transmitted to Eurostat a few months later. These files were then inserted into the NRME.

32. The Eurostat responsible domain managers checked and validated the files which are now released on the Eurostat website next to the corresponding European metadata file (see figure 5).

Figure 5. The dissemination of European and national ESMS files for HICP



For any question on data and metadata, please contact: [EUROPEAN STATISTICAL DATA SUPPORT](#)

1. Contact Top	
1.1. Contact organisation	Eurostat, the statistical office of the European Union
1.2. Contact organisation unit	Unit G6: Price statistics; Purchasing Power Parities
1.5. Contact mail address	2920 Luxembourg LUXEMBOURG
2. Metadata update Top	
2.1. Metadata last certified	28 February 2011

33. In the meantime, more and more national metadata collections in statistical domains are converted to the ESMS standard in using the NRME. This means that the NRME is growing towards an ESS metadata warehouse.

IV. Compliance with the SDMX Standards and Guidelines

34. The Statistical Data and Metadata eXchange⁵ (SDMX) initiative consists of technical and statistical standards and guidelines, together with an IT service infrastructure and IT tools, for a more efficient exchange and sharing of statistical data and metadata.

35. The SDMX standards and guidelines were adopted within the ESS in February 2007 by the Statistical Programme Committee. The Eurostat top management reaffirmed its support to SDMX in 2009 during a Directors' meeting who considered "the use of SDMX compulsory for all new or considerably changed datasets and reference metadata sets [...] before making SDMX compulsory for all domains in Eurostat".

36. Following this, SDMX is now seen as one of the main enablers for implementing the Vision in Eurostat and the ESS. The SDMX standards and guidelines are also fully respected for the collection of the European and national reference metadata.

37. On the one hand, the SDMX statistical standards are followed. Indeed, the 21 concepts composing the ESMS structure belong to the list of 66 cross-domain concepts as defined in the Annex I of the SDMX Content-oriented Guidelines.

38. On the other hand, the development of all the IT tools and applications used in the process of data and metadata collection is based on the SDMX technical standards. Indeed, the NRME has been created according to the SDMX Technical Standards 2.0 and use the SDMX terminology and vocabulary. This concerns also the development of other tools like EMIS but also the Euro-SDMX Registry.

V. A new legislative approach: from stove-pipe legislation to cross-cutting legislation

39. Keeping in mind the objective of integration of the statistical business processes, a re-adaptation of the legislation system for European statistics also appears as a necessity.

40. This was also one main topic discussed at the last ESS Committee in May 2011 where discussions raised about "a comprehensive legislation policy for the ESS."

41. In this context, the Commission Recommendation 2009/498/EC stimulating the harmonisation of reference metadata exchange within the ESS through the use of a common metadata structure can be seen a first step complying with this new legislative policy.

42. Nevertheless, this legal act concerns only the production and exchange of reference metadata within the ESS and additional actions should be undertaken on a more general level for legislating generic parts of the ESS statistical business processes.

43. Therefore, Eurostat proposes a new generation of cross-cutting legislation as additional support towards the further harmonisation and integration of statistical business processes.

44. In the framework of this new legislative strategy, Eurostat drafted two cross-cutting legal acts.

45. The first one is a framework EP/Council Regulation on the use of harmonised metadata within the ESS. This regulation would be accompanied by implementing measures (such as Commission Regulations) defining the ESS reference and structural metadata to be used within the ESS.

46. The second legislation is a cross-cutting EP/Council Regulation harmonising data and metadata exchange processes and exchange standards. The uses of the Eurostat Single Entry Point (eDamis) and of SDMX as standards for the exchange of aggregated data are defined as to be used within the ESS.

⁵ www.sdmx.org

47. These two cross-cutting legislations are complemented by framework legislations for broader statistical domains (such as business statistics).

VI. Conclusions

48. The ESS Vision (based on the Commission Communication No 404/2009) aims at improving the production methods of European statistics and their efficiency within the European Statistical System. The standardisation of reference metadata and the use of the ESS standards are seen as one answer for implementing this Vision with the effect of also reducing costs and burden across the ESS.

49. The standard metadata structures (such as the ESMS or the ESQRS) are increasingly used within the ESS for the metadata production, exchange and dissemination. The implementation of these metadata standards is facilitated through the use of the web-application, the National Reference Metadata Editor, offered to NSI's for metadata production and exchange in the ESS.

50. The integration and the harmonisation of the metadata production and dissemination in the ESS can't be efficient without the use of common technical and statistical standards. SDMX proposes the appropriate answers and plays an essential role for improving the integration and standardisation of business processes within the ESS.

51. Furthermore, an adaptation of the ESS legislative strategy based on horizontal cross-cutting regulations was discussed at the last ESS Committee in May 2011. Eurostat has already drafted two of these cross-cutting regulations concerning the metadata harmonisation and data/metadata exchange which are fully in the streamline of this new strategy. These two draft legislations will further stimulate the statistical business process integration within the ESS.