

**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, 3-5 April 2006)

Topic (ii): Metadata Concepts, Standards, Models and Registries

**Using the MCV terminology for mapping metadata from different institutions:
the case of Eurostat and OECD**

Supporting Paper

Submitted by Eurostat and OECD¹

I. INTRODUCTION

1. This paper highlights a number of key areas where work is currently taking place, particularly within the seven SDMX sponsoring international organisations², to facilitate the standardisation of metadata concepts used by them, and for establishing stronger coordination of metadata requirements by international organisations so that comparable data and metadata can be made available by the relevant providers more easily, reducing redundancies and minimising reporting effort. These tasks entail development of metadata standards in both the area of IT standards (such as the use of common XML formats and tools) and in what the authors refer to as “metadata content guidelines”.

2. The paper illustrates this through a brief presentation of Eurostat and OECD work utilising a small set of standardised concepts envisaged by the SDMX initiative in the development of corporate metadata facilities by both organisations.

II. TERMINOLOGY IS IMPORTANT

3. The Metadata Common Vocabulary (MCV for short) is one of the key projects launched at the very beginning of the SDMX initiative, in 2001-2002, with the aim of developing a common understanding of standard metadata concepts used by statisticians in documenting the collection, processing, storage and dissemination of statistical data. The immediate objective was the development of a glossary of those standard concepts, whose definitions were consistent with existing, relevant international statistical

¹ Prepared by Marco Pellegrino (Eurostat) and Denis Ward (OECD)

² Bank for International Settlements (BIS), European Central Bank (ECB), Eurostat, International Monetary Fund (IMF), OECD, United Nations Statistical Division (UNSD) and World Bank

guidelines and recommendations, with the terminology used within international organizations, national agencies and, to the extent possible, in other related projects to develop international standards.

4. In general terms, the SDMX initiative aims at increasing the efficiency in the exchange of data and metadata within the range of activities of the seven sponsoring international organisations while minimising the reporting burden of national agencies. Any exchange of information can happen within the framework of a bilateral or multilateral exchange between parties, or with the placement of data and metadata on a location that can be accessed by all partners. In both cases, but particularly when the information is shared over the web, there is an essential need for users to understand the nature (and any limitation of use) of the statistics being exchanged. The SDMX standards intend to ensure that appropriate metadata always accompanies the data: for this reason, standards for metadata exchange are extremely important in the SDMX context. The need for a standardisation of metadata terminology is evident if one considers how many times – in actual data and metadata transmissions – the same metadata item is referred to by different names or, conversely, how many times the same name is associated with different concepts.

5. One of the assumptions made at the commencement of the development of the MCV (which flowed from initial Eurostat-OECD efforts on the joint development of statistical glossaries) was that the development and more importantly, the adoption, of a worldwide general format for metadata management by a large number of international organisations and national agencies, for all statistical domains, is not a realistic goal in the foreseeable future. Within this assumption, the MCV focused on a system of definitions for discrete metadata concepts (e.g. source, contact, periodicity, timeliness, reference period, coverage or adjustment) which can be used for any statistical domain and independently from any general model or list of metadata concepts developed by any individual organisation. The Vocabulary is only concerned with the elaboration of these building blocks, subject to ISO terminology standards, easily understandable and re-usable. Agreement on a basic common vocabulary of metadata concepts still provides each agency responsible for compiling metadata with the flexibility of deriving a variety of specific formats and models according to its specific needs. The list of terms and associated definitions simply provides a common language applicable across domains.

III. CURRENT MCV STATUS

A. The MCV structure

6. As mentioned above, the MCV builds on work already undertaken by several organisations and standardisation bodies, rather than confusing the situation by the development of a whole new set of definitions. Where standard definitions were not available or not satisfactory, suitable national definitions have been considered or new ones have been entered.

7. The MCV covers the following conceptual items: general metadata terms (mostly derived from ISO/IEC 11179 and UN or UN/ECE sources); metadata describing statistical methodologies (classifications, data collection, data editing,...), metadata for assessing quality and a selection of terms referring to data / metadata exchange and SDMX terminology (including previous GESMES/TS terms). The opportunity of having one single entry point for accessing a variety of terms, sometimes not available or hard to find on the Internet, is a clear value added. The MCV glossary is readily available on the web through extensive statistical glossary databases such as *CODED* (Eurostat concepts and definitions database) and the *OECD Glossary of Statistical Terms*. Extractions will soon be available in SDMX-compliant XML format to enable more efficient re-use and sharing of content within and outside SDMX boundaries.

8. The MCV is not intended to cover the whole range of statistical terminology, as this area is already covered by other general glossary databases or subject (classifications, data editing) / domain-specific glossaries (such as prices, national accounts, merchandise trade, etc.). The specific function of the MCV is to contain all terms which are normally used for building and understanding metadata systems. A metadata glossary is necessarily linked to a series of other subject-specific glossaries or to more universal statistical glossaries. The insertion within the MCV of some definitions derived from these other glossaries should not be seen as a redundancy, but as a means of resolving the complex and interdisciplinary nature of metadata.

9. The current MCV draft consists of 384 terms (see Annex 1 for the list of entries). It presents the following "fields": term, definition, source, context and related terms. The "context" field is used extensively throughout the glossary, sometimes to provide additional explanations, other times to highlight peculiarities in how a certain definition is applied within a certain domain or geographical context.

10. The MCV also plays an important standardisation role. The SDMX authors have sometimes chosen to present several context explanations for the same term, always quoting in detail the respective source, to show both their peculiarities and, at the same time, to highlight the possibilities for convergence. This is the case for some "quality assessment" items: users can live with different quality frameworks or different meta-models, as long as each concept is well identified, defined and known. In other words, transparency is a prerequisite for a correct interpretation (and for convergence) of any statistical framework³.

B. THE 2006 UPDATE (INTRODUCING SDMX CROSS-DOMAIN CONCEPTS)

11. The SDMX initiative recently delivered a set of draft "content-oriented guidelines" which have been posted on the web (at <http://www.sdmx.org>) for public comment. These guidelines contain recommendations for creating interoperable data and metadata sets using SDMX standards. The work, focused on the harmonisation of a limited range (i.e. 26) of high-level concepts common to a large number of statistical domains, and is aimed at encouraging the exchange of comparable statistical information both between international organisations and between national agencies and international organisations.

12. The SDMX content package also includes a newly revised version of the MCV, where high-level concepts (see Annex 2) have been taken into account and referred to the nearest – sometimes broader – term. In many instances, the "context" field of the MCV has been updated to document the use recommended in order to be SDMX-compliant.

13. The content package emphasises the identification of reference⁴ metadata concepts and subject-matter domains to test SDMX standards. In this context, the MCV plays an important role in providing the common set of terms and definitions that can be used to describe the data. But the standardisation also includes, where appropriate, the representation of concepts with code lists and the identification of the role they play within data and metadata structures.

14. Further work will be undertaken at the end of the current public consultation process to produce a consolidated version of the MCV to ensure that all definitions in the Vocabulary are in line with the agreed list of cross-domain concepts, to improve the content / wording of MCV definitions, and to make available the SDMX-ML version of the glossary.

IV. MAPPING REFERENCE METADATA TO MCV CONCEPTS

15. As outlined below, Eurostat and the OECD currently use the MCV to ensure clarity and terminological consistency within both organisations' respective metadata repositories (Eurostat free dissemination, OECD MetaStore). The use of standard definitions taken from the MCV is similarly encouraged within metadata-related projects and activities of other international organisations and national agencies.

16. The availability of a web repository of standard metadata definitions, available for all Internet users, is also a unique chance for creating a common understanding across countries, for instance across European Union or OECD Member countries.

³ The possible convergence of the quality frameworks of international organisations is currently being investigated by a task force established in 2005 by the Co-ordinating Committee for Statistical Activities (CCSA) whose work will be discussed at the Q2006 conference in Newport, Wales, on 27-28 April 2006.

⁴ In SDMX, "reference" metadata are metadata describing the contents and the quality of the statistical data, normally including "conceptual" metadata, describing the concepts used and their practical implementation; "methodological" metadata, describing methods used for the generation of the data (e.g. sampling, collection methods, editing processes); and "quality" metadata, describing the different quality dimensions of the resulting statistics (e.g. timeliness, accuracy). These metadata are often stored in a separate metadata repository and they are referenced from the related data element.

A. Metadata within Eurostat

17. Eurostat has disseminated data free of charge from its web site since October 2004. Since the introduction of the new dissemination policy, actions have been taken for producing, monitoring and finally improving the quality of metadata descriptions according to a standardised template, built on the conceptual elements used by the Special Data Dissemination Standard (SDDS) format. But the original SDDS model, developed by the IMF for macro-economic and financial domains, was not sufficient for Eurostat purposes: the main reason is that IMF presents information received for each Member State dataset without being involved in further processing. Eurostat, on the other hand, processes and disseminates data and quality information for the entire EU: for this reason, users expect to receive not only national information but also an overall assessment from a Eurostat perspective. This also reflected the differences in the Eurostat and IMF quality frameworks which were taken into account when adapting the metadata elements to better serve the free data dissemination.

18. The adoption of a "corporate" standard presented for Eurostat the advantage of providing a clear target against which different national and international formats could be mapped. The template used hosts a variety of information items, common to most metadata formats. The list of items is generic (applicable to all the domains) and it has been developed to enable authors to post information at the level of detail needed for each domain and for a variety of different users; in this context, not all of the metadata items are expected to be populated by text for all statistical domains.

19. The format layout consists of a "Base Page" and a "Summary Methodology". The former covers the most basic and broad metadata - providing a short description of the domain, in terms of its coverage, periodicity, timeliness, legal basis or agreement and other general items - whereas the summary methodology has a more technical and statistical character. In the summary methodology, specific sub-elements (e.g. statistical units, reference period, adjustments and compilation of EU aggregates) were added to the original chapters in order to provide a more detailed statistical structure within the general template.

20. The information currently contained in more than 550 files - publicly available on Internet and covering the whole range of statistical data disseminated - will be managed by a new database system which is going to be operational by the end of 2006 to facilitate the creation, checking, re-use and dissemination of a detailed list of metadata items, by linking metadata to the dissemination tree. In a second stage, a "metadata handler" will offer the possibility of browsing and navigating in different metadata systems (for classifications, concepts and definitions, methodological texts,...) thereby allowing to further enrich explanatory text.

21. While working on the technical infrastructure, Eurostat is therefore currently improving the granularity of the metadata format with the aim of extending the conceptual coverage, in particular for incorporating more elements on quality assessment, according to the criteria identified by the European statistical code of practice. The list of granular concepts (described in Annex 3) is built on the current format used, with some limited extensions on quality elements which are going to be further detailed by the end of this year. The current list is going to be used for testing the possibility of disseminating a good selection of reference metadata with regard to the SDMX implementation activities with European member States⁵.

B. Metadata within the OECD

22. The current situation with respect to metadata in the OECD takes place within the context of the Organisation's decentralised statistical system wherein statistical data collection, storage and dissemination for 30 Member countries (plus a limited number of non-member states) is conducted by a number of units (Directorates) across the OECD. Only limited metadata are stored in databases actually linked to the statistics they describe. Most metadata currently reside in numerous text files that have been used in the preparation of a large number of statistical and other publications produced across the Organisation. In the

⁵ See METIS 2006, WP 27, *Using SDMX Standards for rapid dissemination of short-term indicators on the European economy*, submitted by Eurostat.

absence of a corporate metadata facility and corporate metadata model there is frequent duplication of metadata storage.

23. A key element of the new OECD corporate data environment currently being developed is the MetaStore⁶ facility which will, for the first time, enable users within the Organisation to store their metadata in a corporate environment that can be readily accessed by different in-house users and allow metadata describing common data disseminated by different Directorates to be linked to different outputs in lieu of duplicated collection and storage. MetaStore also has the capability of storing links (publicly available URLs) to metadata maintained both by other international organisations and national agencies, again in lieu of direct collection. It is also equipped with powerful text search and retrieval facilities. Finally, MetaStore is linked to other elements of the OECD corporate data environment such as the primary external data dissemination facility, OECD.stat and the *OECD Glossary of Statistical Terms*.

24. MetaStore provides sufficient flexibility appropriate to the OECD's decentralised statistical environment. In a worst case scenario it could merely be used as a corporate metadata storage facility to store existing duplicated metadata compiled and disseminated by the various Directorates within the OECD. In order to maximise the advantages and potential such a facility provides the OECD has also developed a set of governance practices, etc, regarding the insertion of new metadata in lieu of using existing metadata both within MetaStore and in the repositories of other organisations (refer footnote 5).

25. The MetaStore facility contains a set of 41 metadata items and their related definitions (refer "Child level items" in Annex 4). When developed 18 months ago these were almost completely consistent with the then existing version of the MCV. The mapping between the MetaStore items and the high level concepts contained in the recently released draft SDMX content-oriented guidelines in Annex 3 below shows that by and large this consistency remains though further adjustments / refinements will be made to the MetaStore items once the SDMX standards become more stable following public consultation. As can be seen in Annex 4, there is either a one to one relationship between the OECD and SDMX concepts or a many to one relationship where the OECD concept is more granular (e.g. coverage, statistical processing).

26. It should be emphasised that the MetaStore list of metadata items has been developed in the context of the requirements of an international organisation, specifically the OECD, where the main need is for broad metadata that describes the statistics collected and disseminated by the Organisation. In this context not all of the 41 metadata items are expected to be populated by metadata text for all statistical outputs disseminated by all Directorates across the OECD. For example, the OECD's *Main Economic Indicators* (MEI) uses a subset of around 12 of the MetaStore items. Similarly, not all the draft SDMX cross-domain concepts are used in MetaStore.

V. FACILITATING THE EXCHANGE OF METADATA THROUGH SDMX: HOW NATIONAL AGENCIES CAN BENEFIT FROM THIS

27. Recent developments in SDMX technical standards (now ISO/TS/17369) encouraged the specification of formal rules for formatting data and metadata, so that these can be exchanged, read and processed automatically. The use of standard concepts can be applied to the exchange of data and metadata sets. This involves the use of metadata concepts of generic use, common to a number of domains, such as "periodicity", "timeliness", "data source", "statistical adjustment", or "compilation", as well as other which may be specific to a statistical subject-matter domain. A web-service, using information about web locations of data and metadata, can then navigate, find and automatically process the information.

28. Through the alignment to these standards by both international organisations and national agencies, there is a concrete possibility of setting the requirements for a concept family of metadata to be exchanged and shared among countries and international organisations. Alignment does not necessarily entail the direct adoption of precisely the same concept (or concepts) by each agency. Although such adoption would

⁶ The MetaStore facility is described in more detail in the paper, *Implementation of MetaStore at the OECD* (Penlington and Thygesen), presented in Session 4 (WP25) at the Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS), Geneva, 3-5 April 2006

facilitate the ability to exchange metadata (with the same “content”) between agencies, it would be sufficient for organisations to be able to map the granular concepts developed to meet their own circumstances and needs to the list of high-level cross-domain concepts specified in the recently released SDMX content-oriented guidelines – as envisaged in the mapping between Eurostat – SDMX concepts and OECD MetaStore metadata items - SDMX concepts in Annexes 3 and 4 below.

29. Agreement on a common set of concepts by national agencies and international organisations would represent a significant step forward. In this context, Eurostat and OECD could play a role in interconnecting European and national metadata for a wide range of indicators of common interest.

30. The mechanism for the actual exchange of metadata between organisations is beyond the scope of the current paper. However, the adoption of the common set of concepts envisaged by the SDMX content-oriented guidelines and the accessibility of metadata – based on these concepts – on national websites would facilitate direct access by international organisations in lieu of the current transmission(s) by national agencies of different metadata to different international organisations. Incidentally, this also demands a good coordination between national and international organisations, to make sure that standardised metadata covering a range of common requirements are made available through the web, while additional information is created directly at supranational level, when this is needed to document the data sets which are disseminated.

VI. CONCLUSION

31. Establishing a framework for the identification of a set of cross-domain concepts implies a dynamic update of the MCV to reflect the SDMX standards as they evolve through the current and future public consultation process. The metadata concepts and perhaps more granular items falling within the core set of terms will be revised and the MCV will need to include new terms and refine existing definitions. In general, as metadata concepts will be subject to periodic revision and supplementation, the MCV will never be considered as complete or final. On the other hand, agreement on a common terminology and on a target set of metadata items allows a simplification of metadata production, the set up of synergies and a reduction of double work for national and international organisations. This also reduces the risk of delivering inconsistent and overlapping information to the users

32. Annexes 2, 3 and 4 to the present paper provide an example of how two organisations could map their internal systems of metadata items to a set of common high-level concepts, and the list of concepts to the standard terminology embedded in the MCV. It would be interesting to know about similar exercises conducted by national institutes and other organisations. Agreement on a limited number of metadata concepts and their mapping to (perhaps more granular) sets of concepts developed by national agencies and international organisations to meet their own particular needs would provide an opportunity for setting the boundaries and contents of a more efficient exchange of metadata. This would also help achieve a reduction of effort at the national level.

LIST OF TERMS DEFINED WITHIN THE METADATA COMMON VOCABULARY (MARCH 2006)

1. Accessibility
2. Accounting basis
3. Accounting conventions
4. Accuracy
5. Adjustment
6. Adjustment Methods
7. Administered item
8. Administration record
9. Administrative data
10. Administrative data collection
11. Administrative source
12. Agency
13. Aggregation
14. Aggregation Equation
15. Analytical framework
16. Analytical unit
17. Area sampling
18. Attachment level
19. Attribute
20. Availability
21. Base period
22. Base weight
23. Base year
24. Basic attribute
25. Benchmark
26. Benchmarking
27. Bias
28. Bilateral exchange
29. Break
30. Category
31. Category Scheme
32. Census
33. Chain index
34. Characteristic
35. Clarity
36. Class
37. Classification
38. Classification changes
39. Classification scheme
40. Classification unit
41. Co-ordination of samples
42. Code
43. Code list
44. Coding
45. Coding error
46. Coherence
47. Collection
48. Comparability
49. Compilation
50. Compilation practices
51. Compiling Agency
52. Completeness
53. Computation of lowest level indices
54. Computer Assisted Interviewing, CAI
55. Concept
56. Concept Scheme
57. Conceptual data model
58. Conceptual domain
59. Confidential data
60. Confidentiality
61. Consistency
62. Consolidation
63. Constraint
64. Contact
65. Context
66. Country identifier
67. Coverage
68. Coverage errors
69. Coverage ratio
70. Creation date
71. Cross-domain Concepts
72. Cut-off survey
73. Cut-off threshold
74. Data
75. Data analysis
76. Data attribute
77. Data capture
78. Data checking
79. Data collection
80. Data confrontation
81. Data consumer
82. Data dissemination
83. Data dissemination standards
84. Data editing
85. Data element
86. Data element concept
87. Data element derivation
88. Data exchange
89. Data exchange context
90. Data flow definition
91. Data identifier
92. Data interchange
93. Data item
94. Data model
95. Data presentation
96. Data processing
97. Data provider
98. Data provider series key

99. Data reconciliation
100. Data security
101. Data set
102. Data set identifier
103. Data sharing exchange
104. Data source
105. Data status upon release
106. Data structure definition
107. Datatype
108. Date
109. Date of last change
110. Definition
111. Derivation input
112. Derivation output
113. Derivation rule
114. Derived data element
115. Derived statistic
116. Dimension
117. Dimensionality
118. Disaggregation
119. Disclosure analysis
120. Dissemination format
121. Documentation
122. Domain groups
123. Dublin Core
124. EDIFACT
125. Electronic data interchange (EDI)
126. Entity
127. Error of estimation
128. Error of observation
129. Estimate
130. Estimation
131. Estimator
132. Expected value
133. Expression Node
134. Flag
135. Flow data series
136. Follow-up
137. Footnote
138. Frame
139. Frame error
140. Frequency
141. Gateway
142. Gateway exchange
143. General Data Dissemination System
(GDDS)
144. Geographical coverage
145. GESMES
146. GESMES/CB
147. GESMES/TS
148. GESMES/TS data model
149. Glossary
150. Graphical data editing
151. Grossing/Netting
152. Guidelines
153. Hierarchy
154. Identifier
155. Imputation
156. Index number
157. Information
158. Information system
159. Inlier
160. Institutional framework
161. Institutional sector
162. Institutional unit
163. Integrity
164. Internal access
165. International code designator
166. International statistical standard
167. Interpolation
168. Interviewer error
169. ISO/IEC 11179
170. Item response rate
171. Key (time series or sibling group)
172. Key family
173. Key structure
174. Keyword
175. Language
176. Level
177. Levels of data
178. Longitudinal data
179. Macro editing
180. Maintenance Agency
181. Measure
182. Measurement error
183. Metadata
184. Metadata Attribute
185. Metadata dimension
186. Metadataflow definition
187. Metadata item
188. Metadata layer
189. Metadata object
190. Metadata registry
191. Metadata set
192. Metadata Structure Definition
193. Metamodel
194. Methodological soundness
195. Methodology
196. Micro editing
197. Ministerial commentary
198. Misclassification
199. Missing data
200. Model assumption error
201. Multilateral exchange
202. Name
203. Nature of the basic data
204. Nomenclature
205. Non-probability sample
206. Non-response
207. Non-response bias
208. Non-response error
209. Non-response rate
210. Non-sampling error
211. Not seasonally adjusted series

- 212. Number raised estimation
- 213. Object
- 214. Object class
- 215. Objectives
- 216. Observation
- 217. Observation confidentiality
- 218. Observation unit
- 219. Observation value
- 220. Ontology
- 221. Organisation
- 222. Organisation identifier
- 223. Organisation Role
- 224. Origin
- 225. Out-of-scope units
- 226. Outliers
- 227. Over-coverage
- 228. Period
- 229. Periodicity
- 230. Permissible value
- 231. Permitted value
- 232. Pre-break observation
- 233. Pre-Break Value
- 234. Precision
- 235. Preferred definition
- 236. Prerequisites of quality
- 237. Primary data
- 238. Primary source of statistical data
- 239. Probability sample
- 240. Processing error
- 241. Product
- 242. Professionalism
- 243. Property
- 244. Provider load
- 245. Provision Agreement
- 246. Public disclosure
- 247. Punctuality
- 248. Qualitative data
- 249. Quality
- 250. Quality (Eurostat context)
- 251. Quality (IMF context)
- 252. Quality (OECD context)
- 253. Quality control survey
- 254. Quality differences
- 255. Quality index
- 256. Quantitative data
- 257. Questionnaire
- 258. Questionnaire design
- 259. Ratio estimation
- 260. Recommended use of data
- 261. Record check
- 262. Record-keeping error
- 263. Recording of transactions
- 264. Reference document
- 265. Reference metadata
- 266. Reference period
- 267. Reference time
- 268. Refusal rate
- 269. Register
- 270. Registrar
- 271. Registration
- 272. Registration authority
- 273. Registry item
- 274. Registry metamodel
- 275. Related data reference
- 276. Related metadata reference
- 277. Relationship
- 278. Relative Standard error
- 279. Release calendar
- 280. Relevance
- 281. Reliability
- 282. Reporting unit
- 283. Respondent burden
- 284. Respondent load
- 285. Response errors
- 286. Response rate
- 287. Responsible organization
- 288. Revision policy
- 289. Sample
- 290. Sample design
- 291. Sample size
- 292. Sample survey
- 293. Sampling
- 294. Sampling error
- 295. Sampling fraction
- 296. Sampling frame
- 297. Sampling technique
- 298. Sampling unit
- 299. Schedule
- 300. Scope
- 301. SDMX-EDI
- 302. SDMX-ML
- 303. SDMX Registry
- 304. Seasonal adjustment
- 305. Secondary source of statistical data
- 306. Semantics
- 307. Serviceability
- 308. Sibling group
- 309. Simultaneous release
- 310. Source
- 311. Source data
- 312. Special Data Dissemination Standard SDDS
- 313. Special language
- 314. Standard Classification
- 315. Standard error
- 316. Statistical concept
- 317. Statistical Data and Metadata Exchange
SDMX
- 318. Statistical error
- 319. Statistical indicator
- 320. Statistical macrodata
- 321. Statistical measure
- 322. Statistical message
- 323. Statistical metadata
- 324. Statistical metadata repository

- 325. Statistical metadata system
- 326. Statistical metainformation
- 327. Statistical metainformation system
- 328. Statistical methodology
- 329. Statistical microdata
- 330. Statistical population
- 331. Statistical processing
- 332. Statistical production
- 333. Statistical standard
- 334. Statistical subject-matter domain
- 335. Statistical unit
- 336. Stewardship
- 337. Stratification
- 338. Structural definition
- 339. Structural metadata
- 340. Structure
- 341. Study domain
- 342. Submission
- 343. Submitting organization
- 344. Supplementary data
- 345. Survey
- 346. Survey data collection
- 347. Survey design
- 348. Syntax
- 349. Target population
- 350. Taxonomy
- 351. Term
- 352. Terminological entry
- 353. Terminological system
- 354. Terminology
- 355. Thesaurus
- 356. Time coverage
- 357. Time of recording
- 358. Time Period
- 359. Time series
- 360. Time series breaks
- 361. Timeliness
- 362. Transparency
- 363. Trend
- 364. Trend estimates
- 365. True value
- 366. Type of data collection
- 367. Under-coverage
- 368. Unit non-response
- 369. Unit of measure
- 370. Unit response rate
- 371. Unit value
- 372. Unit value index
- 373. User needs (for statistics)
- 374. User satisfaction survey
- 375. Validation
- 376. Valuation
- 377. Value domain
- 378. Value item
- 379. Value meaning
- 380. Variable
- 381. Verification
- 382. Weight
- 383. XML
- 384. Year-to-date data

**CURRENT CORRESPONDENCE BETWEEN SDMX CROSS-DOMAIN CONCEPTS FOR
"METADATA STRUCTURE DEFINITIONS" AND MCV TERMS**

Concept	Nearest MCV term
1. Accessibility of documentation	Accessibility
2. Accounting conventions	Accounting conventions
3. Accuracy	Accuracy
4. Classification systems	Classification
5. Comparability/Coherence	Comparability, Coherence
6. Statistical concept	Statistical concept
7. Confidentiality	Confidentiality
8. Contact	Contact
9. Data presentation	Data presentation
10. Date of update	Date of last change
11. Dissemination formats	Dissemination format
12. Frequency and Periodicity	Frequency, Periodicity
13. Institutional framework	Institutional framework
14. Professionalism and ethical standards	Professionalism
15. Quality management (incl. resource management)	Quality
16. Release calendar	Release calendar
17. Relevance	Relevance
18. Revision policy and practice	Revision policy
19. Scope / coverage	Scope, coverage
20. Simultaneous release	Simultaneous release
21. Source data	Source data
22. Statistical processing	Statistical processing
23. Supplementary data	Data dissemination
24. Timeliness and punctuality	Timeliness, Punctuality
25. Transparency	Integrity
26. Validation	Validation

EUROSTAT – SDMX CROSS-DOMAIN CONCEPTS MAPPING

EUROSTAT DISSEMINATION METADATA CONCEPTS		MAPPING TO CURRENT DRAFT OF SDMX CROSS- DOMAIN CONCEPTS
Top level	Child level	
Metadata Update	Last certified without update	Date of update
	Last update of content	Date of update
Contact	Organisation	Contact
	Address	Contact
	Contact name or service	Contact
	e-mail address	Contact
Data coverage	Short description of data domain	Data presentation
	Data breakdown and main variables	Data presentation
	Units of measure	Data presentation
Periodicity	Periodicity of compilation	Frequency and periodicity
	Database frequency	Frequency and periodicity
Timeliness and punctuality	Timeliness	Timeliness and punctuality
	Punctuality	Timeliness and punctuality
Transparency of practices	Legal acts, reporting requirements	Institutional framework
	Rules on confidentiality	Institutional framework
	Internal access	Transparency
	Commentary on the occasion of release	Transparency
Accessibility	Notification of changes in methodology	Transparency
	Release calendar	Release calendar
	Simultaneous release	Simultaneous release
	Dissemination formats	Dissemination formats
	Documentation on methodology	Accessibility of documentation
Quality cross-checks	Related data and quality cross-checks	[No direct concordance]
	References to quality reports	[No direct concordance]
Accuracy and reliability	Overall accuracy assessment	Accuracy
	Quality checks before release	Accuracy
Comparability and coherence	Comparability over time	Comparability and coherence
	Comparability over space	Comparability and coherence
	Comparability with related sources	Comparability and coherence
	Comparability between datasets	Comparability and coherence
	Breaks in time series	Comparability and coherence
Relevance	Rate of available statistics (user needs)	Relevance
	Intended audience and purpose	Relevance
	Supplementary data	Supplementary data
Statistical concepts and classifications	Statistical concept	Statistical concept
	Definition of indicators	Statistical concept
	Classification system	Classification systems
	Conformity with official standards	Classification systems
	Classification coverage	Classification systems
Scope of the data	Reference area / geopolitical entity	Scope/coverage
	Time coverage	Scope/coverage
	Statistical unit	Scope/coverage
	Statistical population	Scope/coverage
Accounting conventions	Reference period	Accounting conventions
	Base period	Accounting conventions
	Basis for recording	Accounting conventions
Nature of basic data	Data source used	Source data
	Type of survey	Source data
	Methods of data collection	Source data
Compilation practices	Compilation	Statistical processing
	Adjustments and weights	Statistical processing
	Data validation	Statistical processing
	Revision policy and practice	Revision policy and practice
Other	Warnings on re-use and limitations	[No direct concordance]

OECD METASTORE – SDMX CROSS-DOMAIN CONCEPTS MAPPING

OECD METASTORE METADATA ITEMS		MAPPING TO CURRENT DRAFT OF SDMX CROSS-DOMAIN CONCEPTS
Top level	Child level	
Source	Contact person and organisation	Contact
	Data source(s) used	Source data
	Name of collection / source used	Source data
	Direct source	Source data
	Source Periodicity	Frequency and periodicity
	Source metadata	Accessibility of documentation
	Date last input received from source	Timeliness and punctuality
Data characteristics and collection	Unit of measure used	Accounting convention / basis
	Power code	Accounting convention / basis
	Variables collected	Statistical concept
	Sampling	Source data
	Periodicity	Frequency and periodicity
	Reference period	Timeliness and punctuality
	Base period	Data presentation
	Date last updated	Date of update
	Link to Release calendar	Release calendar
	Contact person	Contact
	Other Data characteristics and collection	[No direct concordance]
Statistical population and <u>scope of the data</u>	Statistical population	Scope / coverage
	Geographic coverage	Scope / coverage
	Sector coverage	Scope / coverage
	Institutional coverage	Scope / coverage
	Item coverage	Scope / coverage
	Population coverage	Scope / coverage
	Product coverage	Scope / coverage
	Other coverage	Scope / coverage
Statistical concepts and <u>classifications used</u>	Key statistical concepts used	Statistical concept
	Classification(s) used	Classification systems
Manipulation and dissemination	Aggregation & consolidation	Statistical processing
	Estimation	Statistical processing
	Imputation	Statistical processing
	Transformations	Statistical processing
	Validation	Validation
	Index type	Dissemination format
	Weights	Source data
	Seasonal adjustment	Statistical processing
	Other manipulation & adjustments	Statistical processing
	OECD Dissemination format(s)	Dissemination formats
Other aspects	Recommended uses and limitations	Relevance
	Quality comments	[No direct concordance]
	Other comments	[No direct concordance]