



1
2
3
4
5
6
7
8
9
10
11
12
13

SDMX CONTENT-ORIENTED GUIDELINES:

FRAMEWORK

(DRAFT MARCH 2006)



- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36

37 © SDMX 2006
38 [HTTP://WWW.SDMX.ORG/](http://www.sdmx.org/)



39

40 1 INTRODUCTION 4

41 2 SCOPE OF THE CONTENT-ORIENTED GUIDELINES..... 4

42 2.1 Cross-domain Concepts 6

43 2.2 Statistical Subject-Matter Domains..... 6

44 2.3 Metadata Common Vocabulary 6

45 3 DEPENDENCIES ON THE USE OF SDMX TECHNICAL STANDARDS..... 7

46 4 REFERENCES 8

47



48 1 INTRODUCTION

49 The Statistical Data and Metadata Exchange (SDMX) Initiative (<http://www.sdmx.org>)
50 sets technical standards and content oriented guidelines to facilitate the exchange of
51 statistical data and metadata using modern information technology, with an emphasis
52 on aggregated data. Version 1.0 specification of the technical standards has been
53 approved by the International Organization for Standardization (ISO) as Technical
54 Specification; ISO/TS 17369: 2005 SDMX. The Version 2.0 specification (November
55 2005) broadens the framework to support wider coverage of metadata exchange as
56 well as a more fully articulated architecture for data and metadata exchange. Steps will
57 also be taken to bring this work forward within the context of ISO, assuring that SDMX
58 technical standards build on other recognized standards and providing the basis for
59 interoperability with them.

60

61 Unlike the technical specifications in version 1.0 and 2.0, the development of the
62 SDMX content-oriented guidelines outlined below in this document is being undertaken
63 outside the ISO framework.

64

65 The SDMX initiative proposes content-oriented guidelines that are applicable across
66 several statistical subject-matter domains and that represent recommended practices
67 for creating interoperable data and metadata sets using the SDMX technical
68 standards. The content-oriented work is focused on the harmonization of a relatively
69 limited range of specific concepts and terminology that are common to a large number
70 of statistical domains. This harmonisation is necessary to encourage the exchange of
71 comparable statistical information (metadata) and, at this stage, it builds on the
72 experience gained in implementations to date.

73

74 In the early phase of implementation of SDMX standards, the SDMX content-oriented
75 guidelines outlined below are expected to evolve more rapidly than the technical
76 standards, as guidelines-providing agencies and bodies gain experience in the
77 implementation. Also, the SDMX content-oriented guidelines may need to respond to
78 evolving requirements within more statistical domains.

79

80 In addition to proposing cross-domain content-oriented guidelines, the SDMX initiative
81 also provides a structure for the development of domain-specific content-oriented
82 guidelines. Within that framework, recognized international standards-setting agencies
83 and bodies involved in each statistical domain will play an important role in developing
84 domain-specific content-oriented guidelines and their related terminologies.

85

86 2 SCOPE OF THE CONTENT-ORIENTED GUIDELINES

87 This is the first draft release of the SDMX Content-Oriented Guidelines, consisting of:

88

- 89 • *SDMX Content-Oriented Guidelines: Cross-Domain Concepts*
- 90 • *SDMX Content-Oriented Guidelines: Statistical Subject-Matter Domains*
- 91 • *SDMX Content-Oriented Guidelines: Metadata Common Vocabulary*

92

93 It is the intent of these guidelines to establish practices in the use of terminology, the
94 structuring of data and metadata sets, and the classification of data and metadata to
95 support the exchange of data and metadata. These guidelines are complementary to
96 the use of the SDMX Technical Standards. The technical standards may be used



97 independently of the content guidelines, but the content guidelines are designed to
98 work within the specified SDMX technical framework to produce maximum
99 interoperability in the exchange of data and metadata. Thus, the greatest benefits are
100 achieved by using both the framework outlined in the SDMX Technical Standards, and
101 the content guidelines specified here.

102
103 The intent of the SDMX content-oriented guidelines is to encourage reuse and
104 harmonisation where possible across statistical domains in the three following areas,
105 which also define the scope of the guidelines:

- 106
107 (1) Concepts as described in the “Cross-Domain Concepts” guideline. What is
108 presented is a small set of agreed concepts which comes from a very limited
109 initial scope, largely focusing on economic statistics for data-related concepts,
110 and on quality frameworks for metadata-related concepts. There are some
111 additional concepts addressed, but these are the main areas of harmonization.
112 In future, it is anticipated that this scope will rapidly grow to encompass other
113 domains. Having a limited initial guideline was seen as important in starting the
114 process of cross-domain harmonization, however.
- 115
116 (2) Classification of domains as described in the “Statistical Subject-Matter
117 Domains” guideline. What is presented here uses the work of the United
118 Nations Economic Commission for Europe (UNECE) to produce a high-level
119 classification of statistical areas. This classification provides a starting point at
120 the high level for the organization of exchanged statistical data and metadata.
- 121
122 (3) Terminology as described in the “Metadata Common Vocabulary” guideline.
123 The terms presented in the MCV are in many cases taken from other sources,
124 or are harmonised terms used in the SDMX Technical Specifications.
125 Terminology is a broad field, and much effort has been expended over the
126 years in this area. The scope of MCV is the use of a standard terminology in
127 the statistical metadata management across domain boundaries. The MCV, like
128 the other content guidelines, is seen as a living document which will continue to
129 change and grow over time.

130
131
132 Statistical domains cover a very broad field of activities, often with very domain-specific
133 areas of interest. Thus, there will always be domain-specific items at each of the three
134 areas covered by the SDMX Content-Oriented Guidelines. It is not the intent of these
135 guidelines to harmonize everything within all statistical domains. The guidelines
136 provide harmonization where possible, across domain boundaries. Thus, for the
137 instances where various domains use slightly different concepts, or classifications, or
138 terms for the same thing, the SDMX Content-Oriented Guidelines intend to provide a
139 single, harmonized concept, classification, or term to use when exchanging data and
140 metadata across domain boundaries.

141
142 It is important to understand what is meant by the term "cross-domain", as this appears
143 in several places within these guidelines. By "cross-domain", the guidelines indicate
144 that something is used in several domains in a materially similar form. If it meets this
145 criterion, it is considered "cross-domain". This distinction is important, because of the
146 process by which these guidelines are intended to be created and maintained.
147 Identifying all concepts, classifications, and terms which are potentially "cross-domain"
148 according to this definition is a never-ending task. As statistical domains change and

149 expand, new terms and concepts and classifications may come into existence. Rather
150 than attempt to solve this problem in a single effort, the content guidelines are seen as
151 an on-going process which will grow in scope over time.
152

153 **2.1 Cross-domain Concepts**

154 The list of cross-domain concepts, in the SDMX framework, contains and defines
155 metadata concepts relevant to several statistical domains. These concepts are used in
156 SDMX exchange structures; the data structure definition (for data exchange) and the
157 metadata structure definition (for metadata exchange). SDMX recommends use of
158 these concepts whenever feasible to promote re-usability and exchange of statistical
159 information between organizations.
160

161 **2.2 Statistical Subject-Matter Domains**

162 The list of Statistical Subject-Matter Domains is a list covering the breadth of statistical
163 information for the purpose of organizing widespread statistical exchange and
164 categorization. It acts as a standard scheme against which the subject-matter domains
165 of various counterparties can be mapped, to facilitate interoperable data and metadata
166 exchange.
167

168 The list also allows an organization of corresponding “domain groups” comprising
169 international agencies and national agencies working together to develop statistical
170 guidelines and recommendations in a specific domain. In the context of SDMX, each
171 domain group could define standard data structure definitions, concepts, etc. within its
172 individual domain. Such groups already exist within the international community and
173 SDMX would use the statistical subject-matter domains list to focus the energies of
174 these groups on producing any required domain-specific content-oriented guidelines
175 which would be useful to support the implementation of SDMX-conformant technical
176 systems within each domain.
177

178 The SDMX statistical subject-matter domains list will align directly with the UNECE
179 Classification of International Statistical activities and be subject to adjustment in future
180 as the UN classification is revised.

181 **2.3 Metadata Common Vocabulary**

182 The Metadata Common Vocabulary (MCV) is an SDMX repository which contains
183 concepts (and related definitions) to which terminology used in structural and reference
184 metadata of international organisations and national data producing agencies may be
185 mapped.
186

187 The MCV covers a selected range of metadata concepts:

- 189 (1) General metadata concepts, mostly derived from ISO, UNECE and UN
190 documents, useful for providing a general context to metadata management;
191
- 192 (2) Metadata terms describing statistical methodologies (frequency, reference
193 period, data collection, source, adjustment, etc.);
194
- 195 (3) Metadata for assessing quality (accuracy, timeliness, etc.), and
196



197 (4) Terms referring specifically to data and metadata exchange (terminology
198 from the SDMX information model and from existing data structure definitions,
199 etc.).

200

201 More specifically, the MCV provides:

202

203 • ISO/IEC 11179-compliant definitions for a wide range of statistical metadata terms,
204 which may be used directly, or against which other terminology systems may be
205 mapped. This set of terms is inclusive of the terminology used within the SDMX
206 Technical Standards;

207

208 • definitions for cross-domain terms on which the SDMX cross-domain metadata
209 concepts work is built. It plays an important role in providing a repository for the
210 common set of metadata terms and their associated definitions that can be used to
211 describe the collection, processing and dissemination of data; and

212

213 • other terminology used within the SDMX initiative.

214

215 The MCV is not intended to cover the whole range of statistical terminology, as this
216 area is already covered by other general or domain-specific glossaries. The focus of
217 the MCV is largely those terms that are normally used for building and understanding
218 metadata systems.

219

220 Agreement on and updates to the content-oriented guidelines containing and defining
221 SDMX cross-domain metadata concepts imply updating the MCV to reflect these
222 developments.

223

224 **3 DEPENDENCIES ON THE USE OF SDMX** 225 **TECHNICAL STANDARDS**

226 There are dependencies between the SDMX technical standards and the content-
227 oriented guidelines. The technical standards provide a framework within which the
228 content-oriented guidelines are meaningful, although they also may be useful outside
229 the SDMX technical framework. However, the difficult task of agreement and
230 harmonization of concepts will be facilitated by having a known technical framework
231 within which this process takes place. The technical standards provide "ground rules"
232 for the functioning of systems which helps to define the characteristics of meaningful
233 concepts (metadata) harmonization.

234

235 The content-oriented guidelines outlined in this document may be used in addition to
236 the SDMX technical standards. It is possible to be conformant with the SDMX technical
237 standards (ISO/TS 17369:2005 SDMX and version 2.0 SDMX technical specification)
238 and yet not be conformant with the content-oriented guidelines.

239

240 Typically SDMX-ML is used to express the structural metadata which they standardize.
241 To assist in the use of these guidelines, SDMX-ML files are provided as appendices to
242 the appropriate documents. Note that the use of these SDMX-ML expressions is not
243 sufficient for conformance – it merely supports it.

244



245 **4 REFERENCES**

246

247 *SDMX Content-Oriented Guidelines: Cross-Domain Concepts*, SDMX, March 2006.

248

249 *SDMX Content-Oriented Guidelines: Statistical Subject-Matter Domains*, SDMX, March
250 2006.

251

252 *SDMX Content-Oriented Guidelines: Metadata Common Vocabulary*, SDMX, March
253 2006.

254

255 *ISO/TS 17369:2005 Statistical Data and Metadata Exchange (SDMX), version 1.0*,
256 ISO, April, 2005.

257

258 *SDMX Technical Standards, version 2.0*, SDMX, November, 2005.