

SDMX CONTENT-ORIENTED GUIDELINES

ANNEX 2: CROSS-DOMAIN CODE LISTS

DRAFT - 4 FEBRUARY 2008

1. COMMONLY USED CROSS-DOMAIN CODE LISTS

Users of statistics wish to deal with statistics and metadata that are accessible, ideally, in a similar manner. In this sense, also the use of common code lists, to the extent possible, facilitate users to work more efficiently. Moreover, where mappings of code lists are needed, the use of (more) harmonised code lists facilitates the maintenance of the mapping systems and of the interfaces delivered to users. Therefore, choices over code list influence efficiency in data sharing.

SDMX is offering some guidance in this area by initially presenting the following eight code lists. These lists contain code values and code descriptions which are pivot values and descriptions to which the code values used by the SDMX sponsoring organisations or by other organisations can be mapped. On the one hand not all pivot codes are used by all organisations. On the other hand, organisations however may dispose of additional code values and code descriptions which are not included in the following tables.

By helping to start the process of providing code list details, it is hoped that input can be received from the current and future community of users of SDMX standards and guidelines to improve the code values and code descriptions and to identify good practices for harmonising code values and code descriptions.

CL_OBS_STATUS_SDMX.....	3
CL_OBS_CONF_SDMX.....	4
CL_DECIMALS_SDMX.....	4
CL_FREQ_SDMX.....	5
CL_SEX_SDMX.....	5
CL_TIME_FORMAT EDI_SDMX.....	6
CL_TIME_FORMAT_XML_SDMX.....	6
CL_UNIT_MULT_SDMX.....	7

A discussion on particular issues and special cases is provided in the end of this document:	
Information notes on code lists and code values	7

CL_OBS_STATUS_SDMX

Purpose: this code list provides coded information about the "status" of an observation (with respect events such as the ones reflected in the codes composing the code list).

Pivot code value	Pivot code description
A	Normal value
B	Break
E	Estimated value
F	Forecast value

H	Missing value; holiday or weekend
L	Missing value; data exist but were not collected
M	Missing value; data cannot exist
P	Provisional value
Q	Missing value; suppressed
S	Strike

CL_OBS_CONF_SDMX

Purpose: this code list provides coded information about the sensitivity and confidentiality status of the data.

Pivot ode value	Pivot code description
F	Free
N	Not for publication, restricted for internal use only
C	Confidential statistical information
D	Secondary confidentiality set by the sender, not for publication
S	Secondary confidentiality set and managed by the receiver, not for publication

CL_DECIMALS_SDMX

Purpose: it provides a list of values showing the number of decimal digits used in the data.

Pivot ode value	Pivot code description
0	Zero
1	One
2	Two
3	Three
4	Four
5	Five
6	Six

7	Seven
8	Eight
9	Nine

CL_FREQ_SDMX

Purpose: it provides a list of values indicating the "frequency" of the data (e.g. monthly) and, thus, indirectly, also implying the type of "time reference" that could be used for identifying the data with respect time.

Pivot code value	Pivot code description
A	Annual
S	Half-yearly, semester
Q	Quarterly
M	Monthly
W	Weekly
B	Business week
D	Daily
H	Hourly
N	Minutely

CL_SEX_SDMX

Pivot code value	Pivot code description
F	Female
M	Male
T	All genders

CL_TIME_FORMAT_EDI_SDMX

Purpose: these codes are used in the SDMX-EDI format (involving EDIFACT syntax) to indicate the type of time references used in the data.

Pivot code value	Pivot code description
203	CCYYMMDDhhmm
102	CCYYMMDD
616	CCYYWW
610	CCYYMM
608	CCYYQ
604	CCYYS
602	CCYY
711	CCYYMMDDCCYYMMDD
716	CCYYWWCCYYWW
710	CCYYMMCCYYMM
708	CCYYQCCYYQ
704	CCYYS CCYYS
702	CCYYCCYY

CL_TIME_FORMAT_XML_SDMX

Purpose: these codes are used in the SDMX-ML format to indicate the type of time references used in the data.

Pivot code value	Pivot code description
P1D	Daily (or Business)
P1M	Monthly
P1Y	Annual
P3M	Quarterly
P6M	Half-yearly
P7D	Weekly
PT1M	Minutely

CL_UNIT_MULT_SDMX

Purpose: it provides code values for indicating the magnitude in the units of measurements.

Pivot code value	Pivot code description
0	Units
1	Tens
2	Hundreds
3	Thousands
4	Tens of thousands
6	Millions
9	Billions
12	Trillions

2. INFORMATION NOTES ON CODE LISTS AND CODE VALUES

General comments

The use of the SDMX technical standards does not impose on user organisations the direct application of the code lists shown above. However, especially for cross-domain concepts, the mapping of the code lists used by individual organisations to the pivot code values and pivot code descriptions is recommended (i.e. CL_FREQ_SDMX, CL_OBS_STATUS_SDMX, CL_OBS_CONF_SDMX, CL_UNIT_MULT_SDMX, CL_DECIMALS_SDMX; also, CL_TIME_FORMAT_XML_SDMX or CL_TIME_FORMAT_EDI_SDMX, if an SDMX-ML or SDMX-EDI message is used respectively; and CL_SEX_SDMX if the corresponding concept is relevant for the data flow).

CL_OBS_STATUS_SDMX

In some cases there may be more than one or even two “events” that may have influenced the value and are relevant in conjunction with an observation (e.g. a “break” in methodology may be accompanied with the fact that an observation is an “estimate”). Thus, the importance of particular events and, thus, the decision over the most proper flag assignment needs to respect a hierarchy (applicable, if more than one event occurs during the period). The importance is provided on the table below (with "break" being the most important event dominating the other ones).

Observation status hierarchy	Relevant in conjunction with...	
	numeric values	missing values
B / break <i>(highest importance)</i>	✓	✓
M / undefined, data cannot exist		✓
L / data not collected		✓
H / holiday or weekend		✓
S / strike	✓	✓
F / forecast value	✓	
E / estimated value	✓	
P / provisional value	✓	
A / normal value	✓	

This approach (choice of only one event - the most important one - to be described) offers a good compromise between simplicity for the user, completeness of provided information and presentational easiness of management on the user interface side.

Nevertheless, if there is a strong need to describe multiple important events related to a period/observation, also multiple pivot codes can be used.

CL_OBS_CONF_SDMX

The values of this code lists are usually attached at the observation level, providing information about the sensitivity or confidentiality of the corresponding observation. A discussion on the specific code values follows below:

F: this is used for observations for which there are no special sensitivity considerations and which can be freely shared. Please note:

- In some institutional environments the term “unclassified” is used in a sense that still denotes implied restrictions in the circulation of information. If this is the case, this organisation may probably consider that “free” (value F) is not the appropriate corresponding tag for this kind of “unclassified” category and it may be that “non-publishable / restricted” (value N) may be more appropriate.
- The focus of this code list is to simply indicate “sensitivity” and not to describe and anticipate all potential permissible uses. Thus, the use of an “F” does not allow safe assumptions with respect the permission to "re-publish" (freely or at a price) the received or accessed information (e.g. on web or paper), especially on a massive and regular basis (legal and copyright constraints may apply). Usually, source organisations provide information and guidance on republishing issues, either on their websites or on their paper publications.

N: this is used to denote observations that are restricted for internal use only within organisations.

C: Confidential statistical information (*primary confidentiality*) due to identifiable respondents.¹ Measures also should be taken to prevent not only direct access, but also indirect deduction or calculation by other users and parties, probably by considering and treating additional observations as “confidential” (*secondary confidentiality management*).

The following two code values, "**D**" and "**S**", are useful when organisations manage secondary statistical confidentiality. Primary confidentiality corresponds to the special confidentiality status ("confidential statistical information") that applies to data that directly or very easily reveal information about individuals (persons or companies) and which is not publicly available. Obviously, this kind of information needs to be strictly protected. Moreover, if a figure is composed, for example, of information provided by only two agents, then one of these parties or a third party could easily deduct the suppressed information of the individual person or firm. Also, accounting identities could possibly be used or, using longer time series, other algorithms and techniques (e.g. regression) to deduct a suppressed “C” observation out of a large dataset. Thus, there is possibly a need to suppress additional information which is not confidential statistical information in the first place or, in general, to take measures (secondary confidentiality management) against indirect deduction of the suppressed “C” information from third parties who might be using the rest of the information that is available.

D: This flag is used by the sender of the data to flag one or more additional observations of the dataset so that the receiver knows that he/she should suppress these observations in subsequent stages of processing (especially, dissemination) in order to prevent third parties to indirectly deduct (e.g. through accounting identities or other formulas) the observations that were genuinely flagged with "C".

S: If senders do not manage the secondary confidentiality in their data and/or there are also other countries' data involved (with the intention to eventually compile a regional-wide aggregate that is going to be published), this flag is used by the receiver to flag additional suppressed observations (within sender's data and/or within the datasets of other senders) in subsequent stages of processing (especially, dissemination) in order to prevent third parties to indirectly deduct the observations that were genuinely flagged with "C" by the sender.

CL_DECIMALS_SDMX

No special comments.

¹ It may also be used in a case where, in agreement with the other partner institutions in a given data sharing context, the strictest possible sensitivity status needs to be assigned and strict protection measures need to be taken (i.e. access by very few carefully selected users).

CL_FREQ_SDMX

B (business frequency): similar to "daily", however there are no observations for Saturday and Sunday (so, neither "missing values" nor "numeric values" should be provided for Saturday and Sunday). This treatment ("business") is one way to deal with such cases, but it is not the only option. Such a time series could alternatively be considered daily ("D"), with missing values in the weekend.

N (minutely frequency): this is also a slightly exceptional case. While N denotes "minutely", usually, there are no observations every minute (so, the frequency is usually "irregular" within a day/days). And though usually observations are more sparse (not every minute), no missing values are given for the minutes when no observations exist. In any case the time stamp determines when an observation is observed.

CL_SEX_SDMX

No special comments.

CL_TIME_FORMAT_EDI_SDMX

The table below illustrates the potential use of the code values with a few examples.

Types for dates and specific periods			
Code	format	explanation	dates and specific periods: examples
203	CCYYMMDDhhmm	(for series with freq. higher than daily) year/month/day/hours/minutes	199511210850 = 21 Nov. 1995 (08:50) date & format code: 199511210850:203
102	CCYYMMDD	(for daily and business series) year/month/day	19951121 = 21 Nov. 1995 date & format code: 19951121:102
616	CCYYWW	(for weekly series) year/ week	199252 = 52nd week of 1992 date & format code: 199252:616
610	CCYYMM	(for monthly series) year/month	199511 = Nov. 1995 date & format code: 199511:610
608	CCYYQ	(for quarterly series) year/quarter	19953 = 1995q3 date & format code: 19953:608
604	CCYYS	(for half yearly series) year/semester	date & format code: 19951:604
602	CCYY	(for annual series) year	date & format code: 1995:602
Types for period ranges			
Code	format	explanation	period ranges: examples
711	CCYYMMDD CCYYMMDD	(for daily and business series) from year/month/day to year/month/day	1992110419930124 = 4Nov92-24Jan1993 date & format code: 1992110419930124:711
716	CCYYWW CCYYWW	(for weekly series) from year/week to year/week	199227199550 = from week no.27 of 1992 to week no.50 of 1995 date & format code: 199227199550:716

710	CCYYMM CCYYMM	(for monthly series) from year/month to year/month	199208199511= from Aug92 to Nov95 <i>date & format code: 199208199511:710</i>
708	CCYYQCCYYQ	(for quarterly series) from year/quarter to year/quarter	1992319954= from 1992q3 to 1995q4 <i>date & format code: 1992319954:708</i>
704	CCYYSCCYYs	(for half yearly series) from half-year to half-year	1995219961= from 1995h2 to 1996h1 <i>date & format code: 1995219961:704</i>
702	CCYYCCYY	(for annual series) from year to year	19951996= from 1995 to 1996 <i>date & format code: 19951996:702</i>

CL_TIME_FORMAT_XML_SDMX

In case of a need to use the “business” frequency in conjunction with a compact format where the concrete dates (one-by-one) are not explicitly mentioned (e.g. thus, providing only the first date or only the first and last date), please contact the SDMX Secretariat for further clarifications.

CL_UNIT_MULT_SDMX

No special comments.