sdmx

Statistical Data and Metadata eXchange

SDMX STATISTICAL GUIDELINES

_____

Global Metadata Structure Definition (MSD) and Concept Scheme

Version 1.0 – 2/9/2016

_____

INTRODUCTION

This document proposes a Global Metadata Structure Definition and Concept Scheme (hereafter, both artefacts are referred to as "Global MSD" apart from when a useful distinction is made between them).

The overall goal of proposing this Global MSD is to make it possible to increase the efficiency of reference metadata exchange between statistical agencies and for dissemination. The quality and timeliness of the reference metadata can be improved and aligned, thereby enabling statistical data to be described more thoroughly in a comparable way, using automated, shared systems.

The Global MSD was built from the list of SDMX cross-domain concepts relating to metadata in the SDMX Glossary[3] as per 1 June 2016. When the SDMX Glossary was compiled, the methodology used was to include all relevant reference metadata concepts that can be unambiguously mapped to classifications used by Eurostat[1], OECD, ECB, and the IMF. The documented mappings between these organisations and the SDMX concepts are available online[2].

The procedure of building the Global MSD was to structure the SDMX cross-domain concepts relating to reference metadata into homogeneous categories, for example: Administrative Information; Scope; Statistical Processing. The categories are simply guides in this document and are not functional concepts.

For each concept, this document includes the definition and concept ID from the SDMX Glossary for ease of reference. To avoid unnecessary duplication, the usage context, related terms and other fields have not been included here. See the SDMX Glossary[3] for those fields.

The SDMX artefacts for the Global MSD and Concept Scheme are available in the Global Registry[4] for references from system implementations.

---

[1] Eurostat's metadata structure SIMS (Single Integrated Metadata Structure) is made of two parts: ESMS (EURO-SDMX Metadata Structure) and ESQRS (ESS Standard for Quality Reports Structure). Only the concepts from the ESMS part are included in this version of the MSD.

[2] <Include link to mapping document when it is placed online>

[3] https://sdmx.org/wp-content/uploads/SDMX_Glossary_Version_1_0_February_2016.docx

[4] <Link to SDMX artefacts in the Global Registry>

Coverage of Metadataflows

This Global MSD is intended for a cross-domain (or cross-Metadataflow) purpose rather than a domain (or Metadataflow) specific purpose; this is because many statistical agencies have already implemented a cross-domain reference metadata classification. Nevertheless, for certain Metadataflows or domains, it may be relevant to create specific MSDs to limit and/or extend the concepts that may be used. For this reason, the reference metadata concepts are also a Concept Scheme which can be derived from in order to create these specific "custom MSDs".

The concepts in the Global MSD are meant to cover all of the types of reference metadata in the source classifications[5]. However, it is likely that a Metadataflow may not use all of the concepts. Therefore, if it is desired to limit the concepts for a Metadataflow, it is highly recommended to use a subset of concepts from the Concept Scheme; the reason being that the same mapping may be used across separate Metadataflows, and the reuse should avoid flow-specific changes to systems.

The following diagram shows the relationship between the Concept Scheme, Global MSD, and derived custom MSDs that are created for specific domains or Metadataflows. It is recommended that all custom MSDs derive from this Concept Scheme if possible.
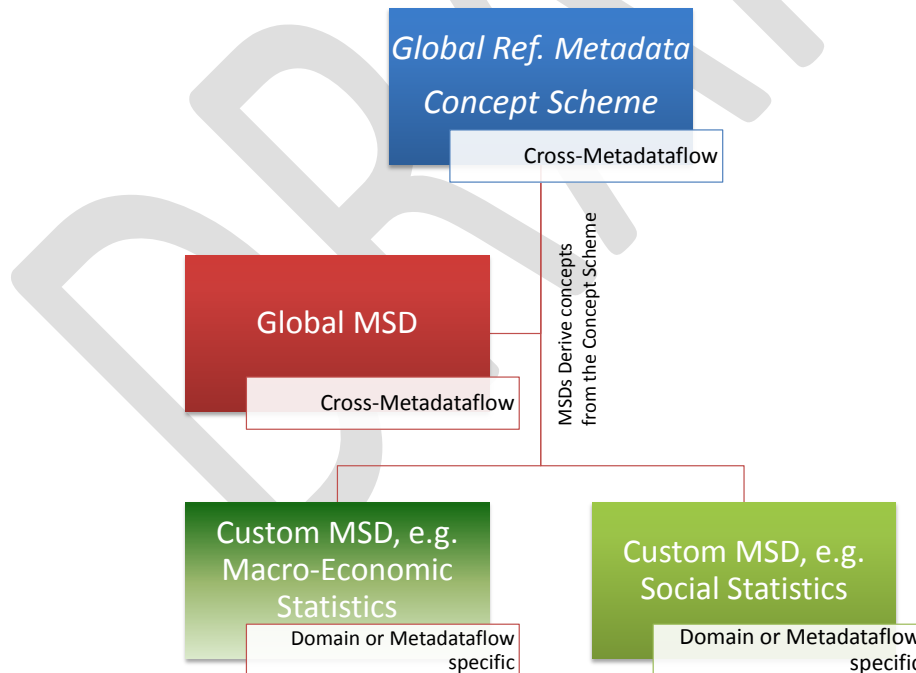
In the case of the need to extend the Concept Scheme concepts, the SDMX Statistical Working Group should be informed to consider including them in a future version of the Concept Scheme, in order to help future alignment of the reference metadata concepts.

---

[5] Reference metadata classifications used by IMF, ECB, OECD, Eurostat.

Implementation of the Global MSD

In order to avoid a large impact on existing systems, it is intended that mappings are created between the existing reference metadata classifications and this Global MSD (the alternative being that statistical agencies change their systems and classifications in line with the MSD – but the cost may be too great).

The mappings between the Global MSD and the reference metadata source classifications[5] are here[6], and it is intended to convert the mappings to SDMX Structure Set artefacts and make them available in the SDMX global registry; this will avoid duplication of the mapping maintenance and provide a central storage. The advantage of using SDMX Structure Sets is that they are part of the SDMX information model, and may be used in system implementation to convert to and from the Global MSD structure and the existing reference metadata classifications.

The following diagram shows an example of the mapping workflow from reporter to receiver:

---
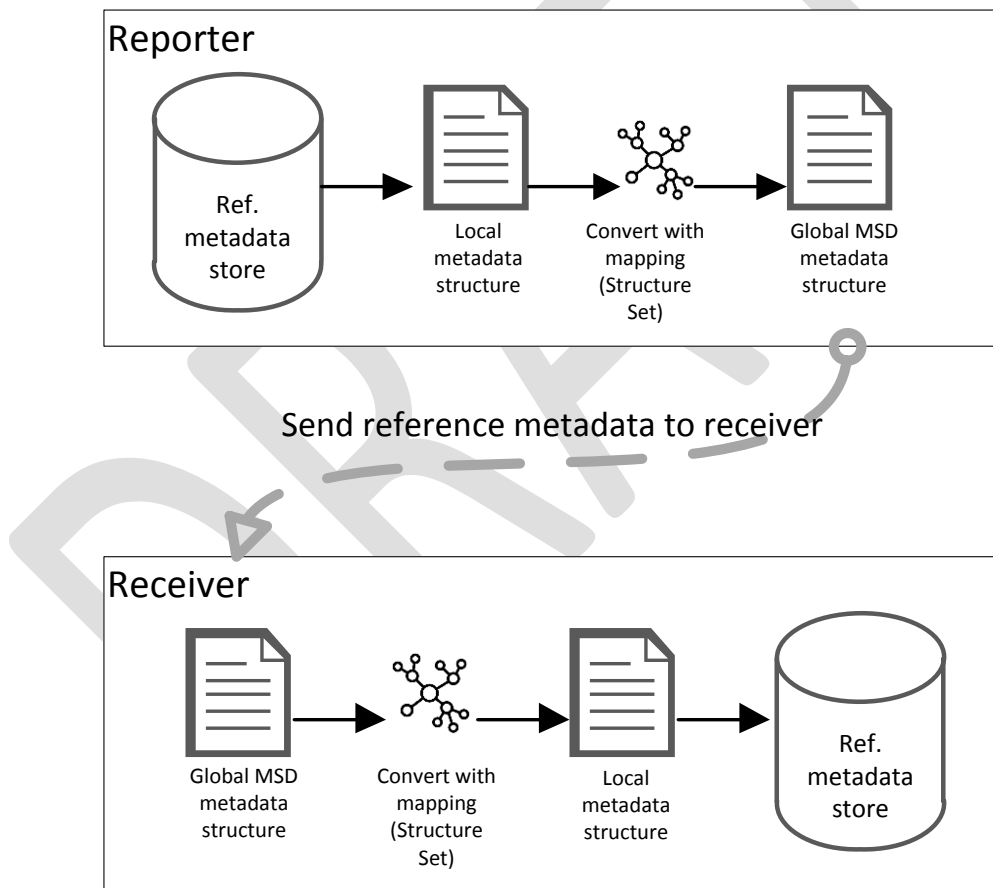
[6] <Include link to mapping document>

CONCEPTS IN THE GLOBAL REFERENCE METADATA CONCEPT SCHEME AND MSD

ADMINISTRATIVE INFORMATION

Name	Definition	Concept ID
Title	Textual label used as identification of a statistical object.	TITLE
Data source	Location or service from where data or metadata can be obtained.	DATA_SOURCE
Contact	Individual or organisational contact points for the data or metadata, including information on how to reach the contact points.	CONTACT
Contact email address	E-mail address of the contact points for the data or metadata.	CONTACT_EMAIL
Contact fax number	Fax number of the contact points for the data or metadata.	CONTACT_FAX
Contact mail address	Postal address of the contact points for the data or metadata.	CONTACT_MAIL
Contact name	Name of the contact points for the data or metadata.	CONTACT_NAME
Contact organisation	Organisation of the contact point(s) for the data or metadata.	CONTACT_ORGANISATION
Contact organisation unit	Addressable subdivision of an organisation.	ORGANISATION_UNIT
Contact person function	Area of technical responsibility of the contact, such as "methodology", "database management" or "dissemination".	CONTACT_FUNCT
Contact phone number	Telephone number of the contact points for the data or metadata.	CONTACT_PHONE

LEGAL AND INSTITUTIONAL ENVIRONMENT

Name	Definition	Concept ID
Institutional mandate	Set of rules or other formal set of instructions assigning responsibility as well as the authority to an organisation for the collection, processing, and dissemination of statistics.	INST_MANDATE
Institutional mandate – data sharing	Arrangements or procedures for data sharing and coordination between data producing agencies.	INST_MAN_SHAR
Institutional mandate - legal acts and other agreements	Legal acts or other formal or informal agreements that assign responsibility as well as the authority to an agency for the collection, processing, and dissemination of statistics.	INST_MAN_LA_OA
Confidentiality	Property of data indicating whether they are subject to dissemination restrictions.	CONF
Confidentiality - data treatment	Rules applied for treating the data set to ensure that private information from individual units cannot be accessed and to prevent unauthorised disclosure.	CONF_DATA_TR
Confidentiality - policy	Legislative measures or other formal procedures which prevent unauthorised disclosure of data that identify a person or economic entity either directly or indirectly.	CONF_POLICY
Professionalism	Standard, skill and ability suitable for producing statistics of good quality.	PROF
Professionalism - code of conduct	Provisions for assuring the qualifications of staff and allowing staff to perform their functions without intervention motivated by non-statistical objectives.	PROF_COND
Professionalism - impartiality	Elements providing assurances that statistics are produced on an impartial basis.	PROF_IMP
Professionalism - methodology	Elements providing assurances that the choices of sources and statistical techniques as well as decisions about dissemination are informed solely by statistical considerations.	PROF_METH
Professionalism - statistical commentary	Elements providing assurances that the statistical entity is entitled to comment on erroneous interpretation and misuse of statistics.	PROF_STAT_COM
Cost and burden	Cost associated with the collection and production of a statistical product, as well as the burden imposed on respondents.	COST_BURDEN

Cost and burden - efficiency management	Cost-benefit analysis, effectiveness of execution of medium term statistical programmes, and ensuring efficient use of resources.	COST_BURDEN_EFF
Cost and burden – resources	Staff, facilities, computing resources, and financing to undertake statistical production.	COST_BURDEN_RES

SCOPE

Name	Definition	Concept ID
Statistical population	Total membership or population or "universe" of a defined class of people, objects or events.	INST_MANDATE
Reference area	Country or geographic area to which the measured statistical phenomenon relates.	REF_AREA
Sector coverage	Main economic or other sectors covered by the statistics.	COVERAGE_SECTOR
Time coverage	Reference metadata element specifying the period of time for which data are provided.	COVERAGE_TIME

METHODOLOGICAL INFORMATION

Data presentation

Name	Definition	Concept ID
Data presentation - Summary description	Total membership or population or "universe" of a defined class of people, objects or events.	INST_MANDATE
Data presentation - Detailed description	Main characteristics of the data set described in an easily understandable manner, referring to the data and indicators disseminated.	DATA_DESCR

Other Methodological Information

Name	Definition	Concept ID
Source data type	Characteristics and components of the raw statistical data used for compiling statistical aggregates.	SOURCE_TYPE
Data collection method	Method applied for gathering data for official statistics.	COLL_METHOD
Reference period	Timespan or point in time to which the measured observation is intended to refer.	REF_PERIOD
Time period	Timespan or point in time to which the observation actually refers.	TIME_PERIOD
Time period - collection	Dates or periods during which the observations have been collected (such as middle, average or end of period) for the target reference period.	TIME_PER_COLLECT
Base period	Period of time used as the base of an index number, or to which a constant series refers.	BASE_PER
Base weight	Weights of a weighting system for an index number computed according to the information relating to the base period instead, for example, of the current period.	BASE_WEIGHT
Frequency of data collection	Time interval at which the source data are collected.	FREQ_COLL
Frequency of observation	Time interval at which observations occur over a given time period.	FREQ
Classification system	Arrangement or division of objects into groups based on characteristics which the objects have in common.	CLASS_SYSTEM
Statistical concepts and definitions	Description of the statistical domain under measure as well as the main variables provided.	STAT_CONC_DEF
Statistical unit	Entity for which information is sought and for which statistics are ultimately compiled.	STAT_UNIT
Unit of measure	Unit in which the data values are expressed.	UNIT_MEASURE
Accounting conventions	Practical procedures, standards and other aspects used when compiling data from diverse sources under a common methodological framework.	ACC_CONV
Valuation	Definition of the price per unit, for goods and services flows and asset stocks.	VALUATION

STATISTICAL PROCESSING

Name	Definition	Concept ID
Data compilation	Operations performed on data to derive new information according to a given set of rules.	DATA_COMP
Adjustment	Set of procedures employed to modify statistical data to enable it to conform to national or international standards or to address data quality differences when compiling specific data sets.	ADJUSTMENT
Price adjustment	Statistical technique used to remove the effects of price influences operating on a data series.	PRICE_ADJUST
Seasonal adjustment	Statistical technique used to remove the effects of seasonal and calendar influences operating on a data series.	SEASONAL_ADJUST
Data validation	Process of monitoring the results of data compilation and ensuring the quality of the statistical results.	DATA_VALIDATION
Data revision	Change in a value of a statistic released to the public.	DATA_REV
Data revision - policy	Policy aimed at ensuring the transparency of disseminated data, whereby preliminary data are compiled that are later revised.	REV_POLICY
Data revision - practice	Information on the data revision practice.	REV_PRACTICE
Data revision - studies	Information about data revision studies and analyses.	REV_STUDY
Imputation	Procedure for entering a value for a specific data item where the response is missing or unusable.	IMPUTATION

QUALITY

Name	Definition	Concept ID
Statistical population	Total membership or population or "universe" of a defined class of people, objects or events.	INST_MANDATE
Quality management	Systems and frameworks in place within an organisation to manage the quality of statistical products and processes.	QUALITY_MGMNT
Quality management - quality assessment	Overall evaluation of data quality, based on standard quality criteria.	QUALITY_ASSMNT
Quality management - quality assurance	Guidelines focusing on quality in general and dealing with quality of statistical programmes, including measures for ensuring the efficient use of resources.	QUALITY_ASSURE
Quality management - quality documentation	Documentation on procedures applied for quality management and quality assessment.	QUALITY_DOC
Accuracy	Closeness of computations or estimates to the unknown exact or true values that the statistics were intended to measure.	ACCURACY
Accuracy - overall	Assessment of accuracy, linked to a certain data set or domain, which is summarising the various components into one single measure.	ACCURACY_OVERALL
Accuracy - non-sampling error	Error in sample estimates which cannot be attributed to sampling fluctuations.	NONSAMPLING_ERR
Accuracy - sampling error	Part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a subset of the population is enumerated.	SAMPLING_ERR
Coherence	Adequacy of statistics to be reliably combined in different ways and for various uses.	COHERENCE
Coherence - cross domain	Extent to which statistics are reconcilable with those obtained through other data sources or statistical domains.	COHER_X_DOM
Coherence - internal	Extent to which statistics are consistent within a given data set.	COHER_INTERNAL
Comparability	Extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics.	COMPARABILITY
Comparability - geographical	Extent to which statistics are comparable between geographical areas.	COMPAR_GEO
Comparability - over time	Extent to which statistics are comparable or	COMPAR_TIME

	reconcilable over time.	
Punctuality	Time lag between the actual delivery of the data and the target date when it should have been delivered.	PUNCTUALITY
Relevance	Degree to which statistical information meets the real or perceived needs of clients.	RELEVANCE
Relevance - user needs	Description of requirements with respect to the statistical output.	USER_NEEDS
Relevance - user satisfaction	Description of how well the disseminated statistics meet the expressed user needs.	USER_SAT
Timeliness	Length of time between data availability and the event or phenomenon they describe	TIMELINESS
Timeliness - source data	Time between the end of a reference period and actual receipt of the data by the compiling agency.	TIME_SOURCE

DISSEMINATION

Name	Definition	Concept ID
Release policy	Rules for disseminating statistical data to interested parties.	REL_POLICY
Release policy - release calendar	Schedule of statistical release dates.	REL_CAL_POLICY
Release policy - release calendar access	Description of how the release calendar can be accessed.	REL_CAL_ACCESS
Release policy - transparency	Statement describing whether and how the release policy is disseminated to the public.	REL_POL_TRA
Release policy - user access	Policy for release of the data to users, scope of dissemination (e.g. to the public, to selected users), how users are informed that the data are being released, and whether the policy determines the dissemination of statistical data to all users.	REL_POL_US_AC
Dissemination format	Media by which statistical data and metadata are disseminated.	DISS_FORMAT
Dissemination format - microdata access	Information on whether micro-data are also disseminated.	MICRO_DAT_ACC
Dissemination format - news release	Regular or ad-hoc press releases linked to the data.	NEWS_REL
Dissemination format - online database	Information about on-line databases in which the disseminated data can be accessed.	ONLINE_DB
Dissemination format - publications	Regular or ad-hoc publications in which the data are made available to the public.	PUBLICATIONS
Dissemination format - other formats	References to the most important other data dissemination done.	DISS_OTHER
Documentation on methodology	Descriptive text and references to methodological documents available.	DOC_METHOD
Documentation on methodology - advance notice	Policy on notifying the public of changes in methodology, indicating whether the public is notified before a methodological change affects disseminated data and, if so, how long before.	ADV_NOTICE

MISCELLANEOUS

Name	Definition	Concept ID
Comment	Descriptive text which can be attached to data or metadata.	COMMENT