BS ISO 17369:2013

BSI Standards Publication

Statistical data and metadata exchange (SDMX)

bsi.

...making excellence a habit.™

National foreword

This British Standard is the UK implementation of ISO 17369:2013.

The UK participation in its preparation was entrusted to Technical Committee ICT/-/4, eBusiness coordination committee.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

Published by BSI Standards Limited 2013

ISBN 978 0 580 78677 8

ICS 35.240.01

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 28 February 2013.

Amendments issued since publication

Date	Text affected

BS ISO 17369:2013

INTERNATIONAL STANDARD

ISO 17369

First edition
2013-01-15

Statistical data and metadata exchange (SDMX)

Données statistiques et échange de métadonnées (SDMX)

Reference number
ISO 17369:2013(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 17369 was prepared by Technical Committee ISO/TC 154, *Processes, data elements and documents in commerce, industry and administration*.

This first edition of ISO 17369 cancels and replaces ISO/TS 17369:2005, which has been technically revised.

Introduction

The Statistical Data and Metadata Exchange (SDMX) initiative (http://www.sdmx.org) sets standards that can facilitate the exchange of statistical data and metadata using modern information technology, with an emphasis on aggregated data.

There are several sections to the SDMX technical specification.

a) The SDMX Framework Document presents the scope and integrated functionality of the concepts and specifications that constitute the SDMX standard.

b) The SDMX Information Model is the information model upon which syntax-specific implementations described in the other sections are based. This is intended for technicians wishing to understand the complete scope of the technical standards in a syntax-neutral form. It includes as an annex a tutorial on UML (Unified Modelling Language).

c) SDMX-ML is the XML format for the exchange of SDMX-structured data and metadata. This document describes the use of the XML syntax in SDMX messages, and is accompanied by a set of XML schemas and sample XML document instances.

d) SDMX-EDI is the UN/EDIFACT format for exchange of SDMX-structured data and metadata. This describes the use of the UN/EDIFACT syntax in SDMX messages.

e) The SDMX Registry Specification provides for a central registry of information about available data and reference metadata, and for a repository containing structural metadata and provisioning information. This specification defines the basic services offered by the SDMX registry: registration of data and metadata; querying for data and metadata; and subscription/notification regarding updates to the registry.

f) The SDMX Technical Notes constitute a guide to help those who wish to use the SDMX specifications. They include notes on the expressive differences of the various messages and syntaxes; versioning; maintenance agencies; the SDMX Registry.

g) Web Services Guidelines constitute a guide for those who wish to implement SDMX using web-services technologies. They place an emphasis on those aspects of web-services technologies (including, but not requiring, an SDMX-conformant registry) which will work regardless of the development environment or platform used to create the web service.

SDMX version 2.0 represented a significant increase in scope, and also provided more complete support in those areas covered in SDMX version 1.0. SDMX version 2.0 is backward-compatible with SDMX version 1.0, so that existing implementations can be easily migrated to conformance with SDMX version 2.0.

SDMX version 2.1 represents a set of changes resulting from several years of implementation experience with SDMX version 2.0. The changes do not represent a major increase in scope or functionality, but do correct some bugs and add functionalities in some cases. Major changes in SDMX-ML include a much stronger alignment of the XML schemas with the information model, to emphasize inheritance and object-oriented features, and increased precision and flexibility in the attachment of metadata reports to specific objects in the SDMX information model.

The idea of backward-compatibility in the standards is based on the information model. In both releases, some non-backward-compatible changes have been made to the SDMX-ML formats. However, the same set of information required to use SDMX version 1.0 will permit the use of the same features in SDMX version 2.0. Thus, a data structure definition (DSD) is easily translated from SDMX version 1.0 to SDMX version 2.0, without requiring any new information regarding structures, etc. There have been no changes to the SDMX-EDI format.

The main changes from SDMX version 1.0 to SDMX version 2.0 can be briefly summarized as follows.

— Reference Metadata: In addition to describing and specifying data structures and formats (along with related structural metadata), SDMX version 2.0 also provides for the exchange of metadata which

is distinct from the structural metadata in SDMX version 1.0. This category includes "reference" metadata (regarding data quality, methodology and similar types: it can be configured by the user to include whatever concepts require reporting); metadata related to data provisioning (release calendar information, description of the data and metadata provided, etc.); and metadata relevant to the exchange of categorization schemes.

— SDMX Registry: Provision is made in SDMX version 2.0 for standard communication with registry services, to support a data-sharing model of statistical exchange. These services include registration of data and metadata, querying of registered data and metadata, and subscription/notification.

— Structural Metadata: The support for exchange of statistical data and related structural metadata has been expanded. Some support is provided for qualitative data; data cube structures are described; hierarchical code lists are supported; relationships between data structures can be expressed, providing support for extensibility of data structures; and the description of functional dependencies within cubes are supported.

The main changes from SDMX version 2.0 to SDMX version 2.1 can be briefly summarized as follows.

— Web-services-oriented changes: Several organizations have been implementing web services applications using SDMX, and these implementations have resulted in several changes to the specifications. Because the nature of SDMX web services could not be anticipated at the time of the original drafting of the specifications, the web services guidelines have been completely re-developed.

— Presentational changes: Much work has gone into using various technologies for the visualization of SDMX data and metadata, and some changes have been proposed as a result, to better leverage this graphical visualization.

— Consistency issues: There have been some areas where the draft specifications were inconsistent in minor ways, and these have been addressed.

— Clarifications in documentation: In some cases it has been identified that the documentation of specific fields within the standard needed clarification and elaboration, and these issues have been addressed.

— Optimization for XML technologies: Implementation has shown that it is possible to better organize the XML schemas for use within common technology development tools which work with XML. These changes are primarily focused on leveraging the object-oriented features of W3C XML Schema to allow for easier processing of SDMX data and metadata.

— Consistency between the SDMX-ML and the SDMX information model: Certain aspects of the XML schemas and UML model have been more closely aligned, to allow for easier comprehension of the SDMX model.

— Technical bugs: Some minor technical bugs have been identified. These bugs have been addressed.

— Support for non-time-series data in the generic format: One area which has been extended is the ability to express non-time-series data as part of the generic data message.

— Simplification of the data structure definition/specific message types: Both time series (SDMX version 2.0 Compact) and non-time series data sets (SDMX version 2.0 cross sectional) use the same underlying structure for a structure-specific formatted message, which is specific to the data structure definition of the data set.

— Simplification and better support for the metadata structure: New use cases have been reported and these are now supported by a re-modelled metadata structure definition.

— Support for partial item schemes such as a code list: The concept of a partial (sub-set) item scheme such as a partial code list for use in exchange scenarios has been introduced.

Statistical data and metadata exchange (SDMX)

1 Scope

This International Standard provides an integrated approach to facilitating Statistical Data and Metadata Exchange (SDMX), enabling interoperable implementations within and between systems concerned with the exchange, reporting and dissemination of statistical data and related metadata.

This International Standard is applicable to any organization that has a need to manage the reporting, exchange and dissemination of its statistical data and related metadata. The information model at the core of this International Standard has been developed to support statistics as collected and used by governmental and supra-national statistical organizations, and this model is also applicable to other organizational contexts involving statistical data and related metadata.

2 Terms, definitions and abbreviated terms

2.1 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

2.1.1
representational state transfer
REST
style of software architecture for distributed hypermedia systems, such as the worldwide web

2.1.2
RESTful web service
RESTful API
simple web service implemented using HTTP and the principles of REST

2.1.3
data set
organized collection of data and associated metadata according to an identified data structure definition

2.1.4
data structure definition
DSD
collection of metadata concepts, their structure and usage when used to collect or disseminate data

2.1.5
metadata set
organized collection of metadata structured according to an identified metadata structure definition

2.1.6
metadata structure definition
collection of metadata concepts, their structure and usage, when used to collect or disseminate reference metadata

2.1.7
reference metadata
metadata describing the contents and the quality of the statistical data

2.1.8
code list
predefined list from which some statistical coded concepts take their values

2.1.9
organization scheme
collection of organizations that play a role in the statistical process lifecycle

EXAMPLE Maintenance agency; data provider; data consumer.

2.1.10
organization
unique framework of authority within which a person or persons act, or are designated to act, towards some purpose

[SOURCE: ISO/IEC 6523-1:1998, 3.1, modified.]

2.1.11
category scheme
hierarchy of categories, which may include any type of useful classification for the organization of data and metadata

2.1.12
category
item at any level within a classification

EXAMPLE Tabulation categories; sections; subsections; divisions; subdivisions; groups; subgroups; classes; subclasses.

2.1.13
categorization
linking of a category to an object, such that sets of objects can be classified

2.1.14
concept scheme
list of concepts that are used in data structure definitions and metadata structure definitions

2.1.15
concept
unit of knowledge created by a unique combination of characteristics

[SOURCE: ISO 1087-1:2000, 3.2.1, modified.]

2.1.16
dataflow definition
descriptive information about the flow of data that providers provide for different reference periods, according to an identified data structure definition

2.1.17
metadataflow definition
descriptive information about the flow of metadata that providers provide for different reference periods, according to an identified metadata structure definition

2.1.18
data provider
organization which produces data or reference metadata

2.1.19
provision agreement
arrangement within which the provider supplies data or metadata

2.1.20
structure set
collection of structural maps that express the semantic equivalence between source and target components

2.1.21
reporting taxonomy
scheme which defines the composition structure of a data report where each component can be described by an independent dataflow definition or metadataflow definition

2.1.22
process
scheme which defines or documents the operations performed on data or metadata

2.1.23
hierarchical code list
organized collection of codes arranged in levels of detail from the broadest to the most detailed level

Note 1 to entry: Each level of the hierarchy is defined in terms of the codes at the next lower level of the hierarchy.

2.2 Abbreviated terms

API	Application Program Interface
GESMES	Generic Statistical Message
HTTP	Hypertext Transfer Protocol
MCV	Metadata Common Vocabulary
OLAP	Online Analytical Processing
SDMX	Statistical Data and Metadata Exchange
SOAP	Simple Object Access Protocol
UML	Unified Modelling Language
UN/EDIFACT	United Nations/Electronic Data Interchange for Administration, Commerce and Transport
WADL	Web Application Description Language
WSDL	Web Service Definition Language

3 Processes and business scope

3.1 Process patterns

SDMX identifies three basic process patterns relevant to the exchange of statistical data and metadata. These can be described as follows.

a) Bilateral exchange: All aspects of the exchange process are agreed between two counterparties, including the mechanism for exchange of data and metadata, the formats, the frequency or schedule, and the mode used for communications regarding the exchange. This is perhaps the most common process pattern.

b) Gateway exchange: Gateway exchanges are an organized set of bilateral exchanges, in which several data and metadata collecting organizations or individuals agree to exchange the collected information with each other in a common format, and according to a common process. This pattern obviates the need for managing multiple bilateral exchanges (in data and metadata collection) across the sharing organizations/individuals. This is also a very common process pattern in the statistical area, where communities of institutions agree on ways to gain efficiencies within the scope of their collective responsibilities.

c) Data-sharing exchange: Data sharing is a mode of exchange where any party which is granted access to the data can obtain it in a known format, along with all relevant metadata, so as to be able to use the data without any prior arrangement with the data provider. Such use can be programmatic in nature, requiring a high degree of standardization both of data and metadata formats and content. It is also typical of a data-sharing scenario that the data be easy to locate, using an online catalogue or registry, although this is not an absolute requirement. In data-sharing mode, the differences between reporting data and disseminating data begin to disappear, i.e. a collector of data simply accesses it and downloads the needed data, much like any other user of the data.

The SDMX standards are designed to support any of the three exchange patterns mentioned. It is possible to use SDMX-EDI or SDMX-ML as a metadata-rich exchange format for bilateral reporting or gateway exchange; it is also possible to use the SDMX formats, web services guidelines and registry specification to fully support a data-sharing pattern. The design of SDMX is such that not all of the offered features need to be used: SDMX standards span the breadth between the use of the standard to provide a simple data and metadata format, through to a fully automated and standardized exchange of data in a data-sharing mode.

The standards specified here specifically support a data-sharing process pattern based on the use of central registry services. Registry services provide visibility into the data and metadata existing within the community, and support the access and use of this data and metadata by providing a set of triggers for automated data and metadata retrieval. The data or metadata itself is not stored in a central registry: these services merely provide a useful set of metadata about the data (and additional metadata) in a known location, so that users/applications can easily locate and obtain whatever data and/or metadata are registered. The use of standards for all data, metadata and the registry services themselves is ubiquitous, permitting a high level of automation within a data-sharing community.

These three process patterns are not mutually exclusive: a single system capable of expressing data and metadata in SDMX-conformant formats could support all three patterns.

In addition to looking at collection and reporting, it is also important to consider the dissemination of data. Data and metadata (no matter how they are exchanged between counterparties in the process of their development and creation) are all eventually supplied to an end user of some type. Often, this is through specific applications inside of institutions. However, more and more frequently, data and metadata are also published on websites in various formats. The dissemination of data and its accompanying metadata on the web is a focus of the SDMX standards. Standards for statistical data and metadata allow improvements in the publication of data: a standard format facilitates the linking of data to metadata, making the data more comprehensible to the end user and making further processing of the data easier.

In discussions of statistical data, there are many aspects of its dissemination which impact data quality: data discovery, ease of use and timeliness. SDMX standards provide support for all of these aspects of data dissemination. Standard data formats promote ease of use and provide links to relevant metadata. The concept of registry services means that data and metadata can more easily be discovered. Timeliness is improved throughout the data lifecycle by increases in efficiency, promoted through the availability of metadata and ease of use.

Even though SDMX is primarily focused on the exchange and dissemination of statistical data and metadata, the standard is also applicable in the context of internal processing of data that are not concerned with the exchange between organizations and users.

3.2 SDMX and process automation

Statistical data and metadata exchanges employ many different automated processes, but some are of more general interest than others. There are some common information technologies that are nearly ubiquitous within information systems today. SDMX aims to provide standards that are most useful for these automated processes and technologies.

Briefly, these can be described as follows.

a) Batch exchange of data and metadata: The transmission of whole or partial databases between counterparties, including incremental updating.

b) Provision of data and metadata on the internet: Internet technology (including its use in private or semi-private TCP/IP networks) is extremely common. This technology includes XML and web services as primary mechanisms for automating data and metadata provision, as well as the more traditional static HTML and database-driven publishing.

c) Generic processes: While many applications and processes are specific to a known set of data and metadata, other types of automated services and processes are designed to handle any type of statistical data and metadata whatsoever. This is particularly true in cases where portal sites and data feeds are made available on the internet.

d) Presentation and transformation of data: In order to make data and metadata useful to consumers, they need to support automated processes that transform them into application-specific processing formats, other standard formats and presentational formats. Although not strictly an aspect of exchange, this type of automated processing represents a set of requirements that needs to be supported if the information exchange between counterparties is itself to be supported.

The SDMX standards specified here are designed to support the requirements of all of these automation processes and technologies.

3.3 Statistical data and metadata

To avoid confusion about which "data" and "metadata" are the intended content of the SDMX formats specified here, a statement of scope is offered. Statistical "data" are sets of often numeric observations which typically have time associated with them. They are associated with a set of metadata values, representing specific concepts, which act as identifiers and descriptors of the data. These metadata values and concepts can be understood as the named dimensions of a multi-dimensional coordinate system, describing what is often called a "cube" of data.

SDMX identifies a standard technique for modelling, expressing and understanding the structure of this multi-dimensional "cube", allowing automated processing of data from a variety of sources. This approach is widely applicable across types of data and attempts to provide the simplest and most easily comprehensible technique that will support the exchange of this broad set of data and related metadata.

The term "metadata" is very broad indeed. A distinction can be made between the following:

— "structural" metadata: those concepts used in the description and identification of statistical data and metadata, and

— "reference" metadata: the larger set of concepts that describe and qualify statistical data sets and processing more generally, and which are often associated not with specific observations or series of data, but with entire collections of data or even the institutions which provide that data.

The SDMX information model provides for the structuring not only of data, but also of "reference" metadata. While these reference metadata structures exist independent of the data and its structural metadata, they are often linked. The SDMX information model provides for the attachment of reference metadata to any part of the data or structural metadata, as well as for the reporting and exchange of the reference metadata and its structural descriptions. This function of the SDMX standards supports many aspects of data quality initiatives, allowing as it does for the exchange of metadata in its broadest sense, of which quality-related metadata are a major part.

Metadata are associated not only with data, but also with the process of providing and managing the flow of data. The SDMX information model provides for a set of metadata concerned with "data provisioning", i.e. metadata which are useful to those who need to understand the content and form of the output of a data provider. Each data provider can describe in standard fashion the content of and dependencies within the data and metadata sets which they produce, and supply information about the scheduling

and mechanism by which their data and metadata are provided. This allows for automation of some validation and control functions, as well as supporting management of data reporting.

SDMX also recognizes the importance of category schemes in organizing and managing the exchange and dissemination of data and metadata. It is possible to classify SDMX artefacts in one or category schemes that facilitate both processing of the data and metadata and data and metadata discovery.

The SDMX standards offer a common model, a choice of syntax and, for XML, a choice of data formats which support the exchange of any type of statistical data meeting the definition above; several optimized formats are specified based on the specific requirements of each implementation, as described below in the SDMX-ML section.

The formal objects in the information model are presented briefly below, but are also discussed in more detail elsewhere in this International Standard.

Figure 1 provides a high level schematic of major artefacts in the SDMX information model.

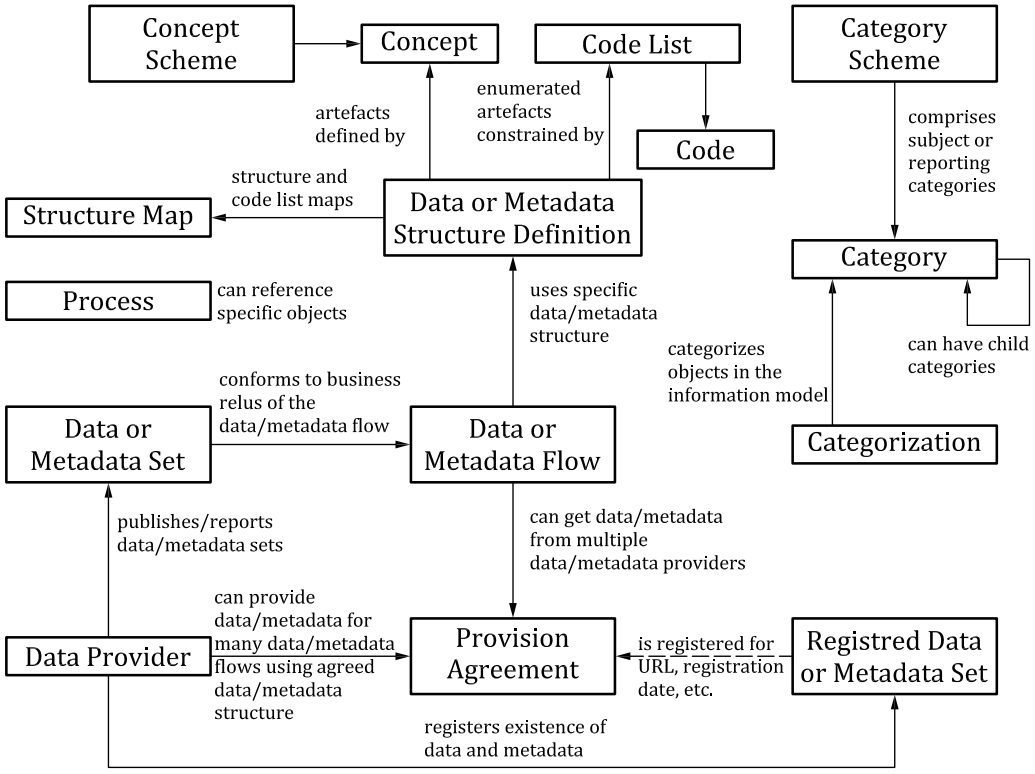

Figure 1 — High level schematic of major artefacts in the SDMX information model

3.4 SDMX view of statistical exchange

3.4.1 General

SDMX version 1.0 (ISO/TS 17369:2005) covered statistical data sets and the metadata related to the structure of these data sets. This scope was useful in supporting the different patterns of statistical exchange (bilateral exchange, gateway exchange and data-sharing), but was not by itself sufficient to support them completely. SDMX version 2.0 and SDMX version 2.1 both provide a much more complete view of statistical exchange, so that the open data-sharing pattern can be fully supported and other patterns of exchange can be more completely automated. In order to produce technical standards that will support this increased scope, the SDMX information model provides a broader set of formal objects which describe the actors, processes and resources within statistical exchanges.

It is important to understand this set of formal objects not only in a technical sense, but also in terms of what these objects represent in the real-world exchange of statistical data and metadata.

SDMX version 1.0 provided for data sets (specific statistical data reported according to a specific structure, for a specific time range) and for data structure definitions (the metadata which describes the structure of statistical data sets). These are important objects in statistical exchanges, and they are retained and enhanced in SDMX version 2.0 and SDMX version 2.1 in a backward-compatible form. A related object in statistical exchanges is the "data flow". This supports the concept of data reporting or dissemination on an ongoing basis: "data flows" can be understood as data sets which are not bounded by time.

SDMX version 2.0 and SDMX version 2.1 additionally provide for the standard the systematic representation of reference metadata. Reference metadata are reported not as an integral part of a data set, but independent of the statistical data. SDMX provides for reference "metadata sets", "metadata structure definitions" and "metadata flows". These objects are very similar to data sets, data structure definitions and data flows, but they concern reference metadata rather than statistical observations. In the same way that data providers may publish statistical data, they may also publish reference metadata. Metadata structural definitions are maintained by agencies in a fashion similar to the way that agencies maintain data structure definitions, the structural definitions of data sets.

The structural definitions of both data and reference metadata associate specific statistical concepts with their representations, whether textual, coded, etc. These concepts are taken from a "concept scheme". A concept scheme groups a set of concepts, provides their definitions and names, and allows for semantic relationships to be expressed when some concepts are specializations of others. It is possible for a single concept scheme to be used both for data structures and for reference metadata structures. An SDMX concept scheme is a collection of concepts. SDMX purposefully does not offer structures for management of concepts and data element concepts such as those described in ISO/IEC 11179-1:2004 and ISO/IEC 11179-3:2003. However, an SDMX concept can be linked by reference to an equivalent construct managed by a concept management system.

Inherent in any statistical exchange (and in many dissemination activities) is the notion of "service level agreement", even if this is not formalized or made explicit. SDMX incorporates this idea in an object termed "provision agreement". Data providers may provide data to many different data flows. Data flows may incorporate data coming from more than one data provider. Provision agreements are the objects that inform which data providers are supplying what data to which data flows. The same is true for metadata flows.

A provision agreement allows for a variety of information to be made available:

a) the schedule by which statistical data or metadata are reported or published,

b) the specific topics about which data or metadata are reported within the theoretically possible set of data (as described by a data structure definition or reference metadata structure definition), and

c) the time period covered by the statistical data and metadata.

This set of information is termed "constraint" in the SDMX information model.

A brief summary of the objects described in the information model includes the following.

— Data set: Data are organized into discrete sets, which include particular observations for a specific period of time. A data set can be understood as a collection of similar data, sharing a structure, which covers a fixed period of time.

— Data structure definition (DSD, also known as key family in SDMX version 2.0): Each data set has a set of structural metadata. These descriptions are referred to in SDMX as data structure definitions, which include information about how concepts are associated with the measures, dimensions and attributes of a data "cube," along with information about the representation of data and related identifying and descriptive (structural) metadata. In SDMX version 2.1, the term "key family" is replaced by "data structure definition" (DSD) both in XML schemas and the information model.

— Code list: Code lists enumerate a set of values to be used in the representation of dimensions, attributes and other structural parts of SDMX. They can be supplemented by other structural metadata which indicates how codes are organized into hierarchies.

— Organization scheme: Organizations and organization structure can be defined in an organization scheme. Specific organization schemes exist for maintenance agency, data provider, data consumer and organization unit.

— Category scheme and categorization: Category schemes are made up of a hierarchy of categories, which in SDMX may include any type of useful classification for the organization of data and metadata. A categorization links a category to an identifiable object. In this way, sets of objects can be categorized. A statistical subject-matter domain scheme is implemented in SDMX as a category scheme.

— Concept scheme: A concept scheme is a maintained list of concepts that are used in data structure definitions and metadata structure definitions. There can be many such concept schemes. A "core" representation of the concept can be specified (e.g. a core code list or other representation, such as "date"). This core representation can be overridden in the data structure definition or metadata structure definition that uses the concept.

— Metadata set: A reference metadata set is a set of information pertaining to an object within the formal SDMX view of statistical exchange: it may describe the schedule on which data are released; it may describe the flow of a single type of data over time; it may describe the quality of data, etc. In SDMX, the creators of reference metadata may take whatever concepts they are concerned with, or obliged to report, and provide a reference metadata set containing that information.

— Metadata structure definition: A reference metadata set also has a set of structural metadata which describes how it is organized. This metadata set identifies what reference metadata concepts are being reported, how these concepts relate to each other (typically as hierarchies), how they may be represented (as free text, as coded values, etc.) and with which formal SDMX object types they are associated.

— Dataflow definition: In SDMX, data sets are reported or disseminated according to a data flow definition. The data flow definition identifies the data structure definition and may be associated with one or more subject matter domains via a categorization (this facilitates the search for data according to organized category schemes). Constraints, in terms of reporting periodicity or sub set of possible keys that are allowed in a data set, may be attached to the data flow definition.

— Metadataflow definition: A metadata flow definition is very similar to a data flow definition, but describes, categorizes and constrains metadata sets.

— Data provider: An organization which produces data or reference metadata is termed a data provider.

— Provision agreement: The set of information which describes the way in which data sets and metadata sets are provided by a data provider. A provision agreement can be constrained in much the same way as a data or metadata flow definition. Thus, a data provider can express the fact that it provides a particular data flow covering a specific set of countries and topics, Importantly, the actual source of registered data or metadata is attached to the provision agreement (in terms of a URL). The term "agreement" is used because this information can be understood as the basis of a "service-level agreement". In SDMX, however, this is informational metadata to support the technical systems, as opposed to any sort of contractual information.

— Constraint: A constraint describes a subset of a data source or metadata source, and may also provide information about scheduled releases of data or metadata. They are associated with data providers, provision agreements, data flows, metadataflows, data structure definitions and metadata structure definitions.

— Structure set: A structure set provides a mechanism for grouping structural metadata together to form a complete description of the relationships between specific, related sets of data and metadata. They can be used to map dimensions and attributes to one another, to map concepts, to map code lists and to map category schemes. They can be used to describe "cubes" of data, even when the data within the cube does not share a common dimensionality.

— Reporting taxonomy: A reporting taxonomy allows an organization to link (possibly in a hierarchical way) a number of cube or data flow definitions, which together form a complete "report" of data or metadata. This supports primary reporting, which often comprises multiple cubes of heterogeneous data, but may also support other collection and reporting functions. It also supports the specification of publications such as a yearbook, in terms of the data or metadata contained in the publication.

— Process: A process provides a way to model statistical processes as a set of interconnected process steps. Although not central to the exchange and dissemination of statistical data and metadata, having a shared description of processing allows for the interoperable exchange and dissemination of reference metadata sets which describe processes-related concepts.

— Hierarchical code list: This supports the specification of code hierarchies. The codes themselves are referenced from the code lists in which they are maintained. The hierarchical code list thus specifies the organization of the codes in one or more hierarchies, but does not define the codes themselves.

3.4.2 Data structuring

SDMX recognizes that statistical data are structured: in SDMX a formal description of this structure is termed a data structure definition. "Data sets" are made up of one or more contained "series" or "sections" of data. Each series or section has a "key" (i.e. values for each of a cluster of concepts, also called "dimensions"), which identifies it, and one or more "observations", which typically combine the time of the observation, and the value of the observation (e.g. measurement). Additionally, metadata may be attached at any level of this structure as descriptive "attributes". Code lists (enumerations) and other patterns for representation of data and metadata are also modelled.

A "cube" is a rich, multi-dimensional construct, which can be viewed along any of its axes (or "dimensions"). In essence, a multi-dimensional matrix is created, with a data point associated with each possible combination of values within the matrix. This is a common technique in such technologies as OLAP and data warehousing.

The view of data in many SDMX formats is primarily as time series, i.e. as a set of observations which are organized around the time dimension, so that each observation occurs progressively through time. There are, however, many types of statistical data which are not typically organized for exchange as time series where data are organized around some other, non-time dimension of the cube (this is often called "cross-sectional" data). SDMX supports a unified format that represents in the data set an organization of the data along any single dimension. In this context, time series is a particular case of the unified format.

Another type of structure commonly found in statistical "cubes" of data is the hierarchical classification, used to describe the points along any of its dimensions (or axes). In SDMX version 1.0, SDMX standards did not provide full support for this functionality. These hierarchical classifications (the hierarchical code list) are supported in SDMX version 2.0 and SDMX version 2.1.

3.4.3 Reference metadata structuring

Metadata structures are based on the idea that concepts can be organized into semantic hierarchies and that these hierarchies can form the basis for the structuring of XML reporting formats.

In addition to the structures of data sets, there is additional information which is also important in managing data exchange and dissemination. Often thought of as "footnote metadata" or "quality metadata", this class of additional information is generally less structured than the data itself. Further, it may be associated not just with a data set, but also with an organization or an exchange process.

To support the structures of this type of metadata, SDMX offers a type of configurable metadata report: "reference metadata". To structure such a report, any statistical concepts needed are defined and then organized into a flat list or hierarchy. The use and representation of each concept is formally described as a metadata attribute: a field within the report structure. The report is then required to have a formally-specified subject, the "target" of the report. Any formally recognized object which exists within the SDMX information model (or any combination of objects) can be defined as the component objects for that class of metadata report. For example, a report structure might specify that a class of metadata

report is associated with a data provider, with a set of metadata attributes describing contact details for the data of that organization.

All information about the metadata attributes and their targets is expressed in SDMX as a metadata structure definition.

As with data structures, the generic format for metadata sets provides a known document structure, while the structure specific format is derived specifically from a metadata structure definition and can perform a higher degree of schema validation.

3.5 SDMX registry services

In order to provide visibility into the large amount of data and metadata which exists within the SDMX model of statistical exchange, a set of registry services is specified in the standard. A "registry" (as understood in web-services terminology) is an application which maintains and stores metadata for querying, and which can be used by any other application in the network with sufficient access privileges. For SDMX, the registry can be understood both as a structural metadata repository and as the index of a distributed database or metadata repository, which is made up of all the data sets and reference metadata sets of the data provider within a statistical community, located across the internet or similar network.

The SDMX registry services are not concerned with the storage of data sets or reference metadata sets. The assumption is that data and reference metadata lives on the sites of its data providers. The SDMX registry services concern themselves with providing visibility of the data and reference metadata, and information needed to access the data and reference metadata. Thus, a registered data set will have its URL available in the registry, but not the data itself. An application which wishes to access that data would query the registry, perhaps by drilling down via a category scheme and dataflow, for the URL of a registered data source, and then retrieve the data directly from the data provider using the query mechanism defined in the SDMX web services guidelines.

SDMX does not require a particular technology implementation of the registry: instead, it specifies the standard interfaces which may be supported by a registry. Thus, users may implement an SDMX-conformant registry in any fashion they choose, so long as the interfaces are supported as specified here. These interfaces are expressed as XML documents and, for query, also as a RESTful interface.

The registry services discussed here can be briefly summarized as follows.

— Maintenance of structural metadata: This registry service allows users with maintenance agency access privileges to submit and modify structural metadata. In this aspect, the registry is acting as a structural metadata repository. However, it is permissible in an SDMX structure to submit just the "stub" of the structural object, such as a code list, and for this stub to reference the actual location from where the metadata can be retrieved, either from a file or a structural metadata resource, such as another registry.

— Registration of data and metadata sources: This registry service allows users with maintenance agency access privileges to inform the registry of the existence and location (for retrieval) of data sources and reference metadata sources. The registry stores metadata about these sources and links it to the structural metadata that give sufficient structural information for an application to process it, or for an application to discover its existence.

— Querying: The registry services have interfaces for querying the metadata contained in a registry, so that applications and users can discover the existence of data sources and reference metadata sources, structural metadata, the providers/agencies associated with those objects and the provider agreements which describe how the data and metadata are made available, and how they are categorized.

— Subscription/notification: It is possible to "subscribe" to specific objects in a registry, so that a notification will be sent to these subscribers whenever the registry objects are updated.

The scope for such a registry and the services it provides has been influenced by the registry services and models in ISO/IEC 11179-3:2003 and ISO 15000-4:2004.

The syntax is specified in SDMX Standards Section 5.

3.6 Web services

Web services allow computer applications to exchange data directly over the internet, essentially allowing modular or distributed computing in a more flexible fashion than ever before. In order to allow web services to function, however, many standards are required:

a) for requesting and supplying data;

b) for expressing the enveloping data which is used to package exchanged data;

c) for describing web services to one another, to allow for easy integration into applications that use other web services as data resources.

SDMX provides guidelines for using these standards in a fashion which will promote interoperability among SDMX web services and allow for the creation of generic client applications that will be able to communicate meaningfully with any SDMX web service which implements these guidelines.

More specifically, the SDMX web services guidelines offer the following.

— An interface (WSDL) for SOAP-based web services: This covers registry submission, and data, metadata and structural metadata query using the SDMX-ML query message.

— An interface (WADL) for RESTful web services: The RESTful API focuses on simplicity. The aim is not to replicate the full semantic richness of the SDMX-ML query message, but to make it simple to perform a limited set of standard queries for both data and structural metadata.

— A list of common error codes: When web services are used, it is necessary to have error codes which can help to explain the situation when problems are encountered. This is introduced in SDMX version 2.1.

The syntax is specified in SDMX Standards Section 7.

4 SDMX information model

The SDMX information model formally describes the objects and object relationships in the SDMX standard. The syntax formats for SDMX are derived directly from the information model and consequently it is possible to build many SDMX applications that are syntax agnostic, i.e. the application is built using an implementation of the information model constructs. The information model is a valuable tool for understanding and using the format specifications and for mapping at the logical level between SDMX and other logical models such as the model for core components in ISO/TS 15000-5:2005.

The SDMX information model is presented using UML, and is also described in prose.

The SDMX information model is specified in SDMX Standards Section 2.

5 SDMX-EDI

The SDMX-EDI format is the UN/EDIFACT GESMES message format conforming to the GESMES/TS version 3.0 implementation guide, as published as a standard of the SDMX initiative.

a) Statistical definitions: An expression of the structural metadata covered by the SDMX information model in a UN/EDIFACT format.

b) Statistical data: Optimized for the batch exchange of large amounts of time series data between counterparties, it allows for extremely compact expression of large whole or partial data sets. Non-

time series data, such as cross-sectional, can be supported if represented as repackaged time series, but there is no direct support for cross-sectional data in this format.

c) Data set list: a list of data sets and their structural metadata.

The SDMX information model provides the constructs which are found in the EDIFACT syntax used for SDMX-EDI, and those found in the XML syntax of SDMX-ML. Since both syntactic implementations reflect the same logical constructs, SDMX-EDI data and structural metadata messages can be transformed into corresponding SDMX-ML formats, and vice-versa. Thus, these standards provide for interoperability between the UN/EDIFACT-based and XML-based systems processing and exchanging statistical data and metadata.

The syntax is specified in SDMX Standards Section 4.

6 SDMX-ML

While the SDMX-EDI format is primarily designed to support batch exchange, SDMX-ML supports a wider range of requirements. XML formats are used for many different types of automated processing and thus need to support more varied processing scenarios. That is why there are several types of messages available as SDMX-ML formats. Each is suited to support a specific set of processing requirements.

a) Structure definition: All SDMX-ML message types share a common XML expression of the metadata needed to understand and process a data set or metadata set, and additional metadata about category schemes and organizations is included. In addition, the structural aspects of data and metadata provision (dataflows and metadataflows) are described using this format.

b) Generic data: All statistical data expressible in SDMX-ML can be marked up according to this data format, in agreement with the contents of a structure definition message. It is designed for any scenario where applications receiving the data need to process it according to a single format. Such applications may need independent access to the structure of the data set before they process it. Data marked up in this format are not particularly compact, but they make easily available all aspects of the data set. This format does not provide strict validation between the data set and its structural definition using a generic XML parser. It supports the transmission of partial data sets (incremental updates) as well as whole data sets. It supports both the time-series and the cross-sectional use cases.

c) Structure-specific data: This format is specific to the data structure definition of the data set (in other terms, it is DSD-specific) and is created by following mappings between the metadata constructs defined in the DSD and the technical specification of the format. It supports the exchange of large data sets in XML format (typically the size of the data set is 50 % of the same data expressed as generic data), provides strict validation of conformance with the DSD using a generic XML parser, and supports the transmission of partial data sets (incremental updates) as well as whole data sets. The structure-specific data format specified in SDMX version 2.1 supports both the time-series and the cross-sectional use cases which were covered by two distinct formats in SDMX version 2.0.

   Many XML tools and technologies have expectations about the functions performed by an XML schema, one of which is a very direct relationship between the XML constructs described in the XML schema and the tagged data in the XML instance. Strong data typing is also considered normal, supporting full validation of the tagged data. These message types are designed to support validation and other expected XML schema functions.

d) Generic metadata: All reference metadata expressible in SDMX-ML format can be marked up according to this schema. It performs only a minimum of validation and is somewhat verbose, but it does support the creation of generic software tools and services for processing reference metadata.

e) Structure-specific metadata: For each metadata structure definition, an XML schema specific to that structure can be created, to perform validation on sets of reported metadata. This structure is less verbose than the generic metadata format and, because the XML mark-up relates directly to the reported concepts, it is appropriate for applications that are designed to process a specific type of metadata report. It is analogous to the structure-specific data format for data in its approach to the use of XML.

f) Query: Data and metadata are often published in databases which are available on the web. Thus, it is necessary to have a standard query document which allows the databases to be queried, and return an SDMX-ML data, reference metadata, or structure message. The query document is an implementation of the SDMX information model for use in web services and database-driven applications, allowing for a standard request to be sent to data providers using these technologies.

g) Registry: All of the possible interactions with the SDMX registry services are supported using SDMX-ML interfaces. All but one of these documents (registry notification) are based on a synchronous exchange of documents, i.e. a "request" message answered by a "response" message. There are two basic types of request: a "Submit", which writes metadata to the registry services, and a "Query", which is used to discover that metadata. Registry interactions provide formats for all types of provisioning metadata, as well as for subscription/notification, structural metadata and data and metadata registration.

Because all of the SDMX-ML formats are implementations of the same information model and all the data and metadata messages are derivable from the structure message which describes a data set or metadata set, it is possible to have standard mappings between each of the similar formats. These mappings can be implemented in generic transformation tools, useful to all SDMX-ML users, and not specific to the DSD of a particular data set or the MSD of a particular metadata set (even though some of the formats they deal with may be). Part of the SDMX-ML package is the set of mappings between the structure-specific data and metadata formats and the structure definition format from which all are derivable.

The syntax is specified in SDMX Standards Section 3A and SDMX Standards Section 3B.

7 Dependencies on SDMX content-oriented guidelines

7.1 General

The technical standards proposed here are designed so that they can be used in conjunction with other SDMX guidelines which are more closely tied to the content and semantics of statistical data exchange. The SDMX information model works equally well with any statistical concept, but to encourage interoperability, it is also necessary to standardize and harmonize the use of specific concepts and terminology. To achieve this goal, SDMX creates and maintains guidelines for cross-domain concepts, terminology and structural definitions. These guidelines can be divided into three main parts, as outlined in 7.2 to 7.4.

7.2 Cross-domain concepts

The SDMX cross-domain concepts specification is a content guideline concerning concepts which are used across statistical domains. The list of concepts in this specification is expected to grow and to be subject to revision as SDMX is used in a growing number of domains. The use of the SDMX cross-domain concepts, where appropriate, provides a framework to further promote interoperability among organizations using the technical standards presented here. The harmonization of statistical concepts includes not only the definitions of the concepts and their names, but also, where appropriate, their representation with standard code lists and the role they play within data structure definitions and metadata structure definitions.

The intent of this guideline is twofold:

— to provide a core set of concepts that can be used to structure statistical data and metadata, in order to promote interoperability between systems ("structural metadata", as described above);

— to promote the exchange of metadata more widely, with a set of harmonized concept names and definitions for other types of metadata ("reference metadata", as defined above).

7.3 Metadata common vocabulary

The metadata common vocabulary (MCV) is an SDMX guideline which provides definition of terms to be used for the comparison and mapping of terminology found in data structure definitions and in other aspects of statistical metadata management. Essentially, it provides definitions for a wide range of statistical terms, which may be used directly, or against which other terminology systems may be mapped. This set of terms is inclusive of the terminology used within the SDMX technical standards.

The MCV provides definitions for terms on which the SDMX cross-domain metadata concepts work is built.

7.4 Statistical subject-matter domains

The statistical subject-matter domains is a listing of the breadth of statistical information for the purposes of organizing widespread statistical exchange and categorization. It acts as a standard scheme against which the categorization schemes of various counterparties can be mapped, in order to facilitate interoperable data and metadata exchange. It serves another useful purpose, however, which is to allow an organization of corresponding "domain groups", each of which could define standard data structure definitions, concepts, etc. within their domains. Such groups already exist within the international community. SDMX would use the statistical subject-matter domains list to facilitate the efforts of these groups to develop the kinds of content standards which could support the interoperation of SDMX-conformant technical systems within and across statistical domains. The organization of the content of such schemes is supported in SDMX as a category scheme.

SDMX statistical subject-matter domains is listed and maintained by the SDMX initiative and is subject to adjustment.

Bibliography

[1] ISO 1087-1:2000, *Terminology work — Vocabulary — Part 1: Theory and application*

[2] ISO/IEC 6523-1:1998, *Information technology — Structure for the identification of organizations and organization parts — Part 1: Identification of organization identification schemes*

[3] ISO 9735:1988, *Electronic data interchange for administration, commerce and transport (EDIFACT) — Application level syntax rules*

[4] ISO/IEC 11179-1:2004, *Information technology — Metadata registries (MDR) — Part 1: Framework*

[5] ISO/IEC 11179-3:2003, *Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes*

[6] ISO 15000-4:2004, *Electronic business eXtensible Markup Language (ebXML) — Part 4: Registry services specification (ebRS)*

[7] ISO/TS 15000-5:2005, *Electronic Business Extensible Markup Language (ebXML) — Part 5: ebXML Core Components Technical Specification, Version 2.01(ebCCTS)*

[8] Extensible Markup Language (XML) 1.0 (Fifth Edition), W3C Recommendation[1)]

[9] SDMX Standards Section 1 — Framework for SDMX technical standards

[10] SDMX Standards Section 2 — Information model: UML conceptual design

[11] SDMX Standards Section 3A — SDMX-ML: Schema and documentation

[12] SDMX Standards Section 3B — SDMX-ML: XML schemas, samples, WADL and WSDL

[13] SDMX Standards Section 4 — SDMX-EDI syntax and documentation

[14] SDMX Standards Section 5 — SDMX registry specification: Logical functionality and logical interfaces

[15] SDMX Standards Section 6 — SDMX technical notes

[16] SDMX Standards Section 7 — Guidelines for the use of web services

1) Available from http://www.w3.org/TR/2008/REC-xml-20081126/

ICS 35.240.01
Price based on 15 pages

This page deliberately left blank

British Standards Institution (BSI)

BSI is the national body responsible for preparing British Standards and other standards-related publications, information and services.

BSI is incorporated by Royal Charter. British Standards and other standardization products are published by BSI Standards Limited.

About us

We bring together business, industry, government, consumers, innovators and others to shape their combined experience and expertise into standards -based solutions.

The knowledge embodied in our standards has been carefully assembled in a dependable format and refined through our open consultation process. Organizations of all sizes and across all sectors choose standards to help them achieve their goals.

Information on standards

We can provide you with the knowledge that your organization needs to succeed. Find out more about British Standards by visiting our website at bsigroup.com/standards or contacting our Customer Services team or Knowledge Centre.

Buying standards

You can buy and download PDF versions of BSI publications, including British and adopted European and international standards, through our website at bsigroup.com/shop, where hard copies can also be purchased.

If you need international and foreign standards from other Standards Development Organizations, hard copies can be ordered from our Customer Services team.

Subscriptions

Our range of subscription services are designed to make using standards easier for you. For further information on our subscription products go to bsigroup.com/subscriptions.

With **British Standards Online (BSOL)** you'll have instant access to over 55,000 British and adopted European and international standards from your desktop. It's available 24/7 and is refreshed daily so you'll always be up to date.

You can keep in touch with standards developments and receive substantial discounts on the purchase price of standards, both in single copy and subscription format, by becoming a **BSI Subscribing Member**.

PLUS is an updating service exclusive to BSI Subscribing Members. You will automatically receive the latest hard copy of your standards when they're revised or replaced.

To find out more about becoming a BSI Subscribing Member and the benefits of membership, please visit bsigroup.com/shop.

With a **Multi-User Network Licence (MUNL)** you are able to host standards publications on your intranet. Licences can cover as few or as many users as you wish. With updates supplied as soon as they're available, you can be sure your documentation is current. For further information, email bsmusales@bsigroup.com.

BSI Group Headquarters

389 Chiswick High Road London W4 4AL UK

Revisions

Our British Standards and other publications are updated by amendment or revision.

We continually improve the quality of our products and services to benefit your business. If you find an inaccuracy or ambiguity within a British Standard or other BSI publication please inform the Knowledge Centre.

Copyright

Useful Contacts:

Customer Services
Tel: +44 845 086 9001
Email (orders): orders@bsigroup.com
Email (enquiries): cservices@bsigroup.com

Subscriptions
Tel: +44 845 086 9001
Email: subscriptions@bsigroup.com

Knowledge Centre
Tel: +44 20 8996 7004
Email: knowledgecentre@bsigroup.com

Copyright & Licensing
Tel: +44 20 8996 7070
Email: copyright@bsigroup.com

bsi.

...making excellence a habit.™