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)

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