TECHNICAL SPECIFICATION

ISO/TS 17369

First edition
2005-12-15

Statistical data and metadata exchange (SDMX)

Données statistiques et échange de métadonnées (SDMX)

Reference number
ISO/TS 17369:2005(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

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/TS 17369 was prepared by the Statistical Data and Metadate Exchange Initiative (as Version 1.0) and was adopted, under a special “fast-track procedure”, by Technical Committee ISO/TC 154, *Processes, data elements and documents in commerce, industry and administration*, in parallel with its approval by the ISO member bodies.

STATISTICAL DATA AND METADATA EXCHANGE INITIATIVE

FRAMEWORK

FOR

SDMX STANDARDS

(VERSION 1.0)

I. 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 this technical specification:

1. The SDMX Information Model - the information model on which syntax-specific implementations described in the other sections are based. This document includes as appendixes a UML tutorial and a tutorial for those who are unfamiliar with key families as a way of describing statistical data structures. This document is not normative.
2. SDMX-EDI - the EDIFACT format for exchange of SDMX-structured data and metadata. This document contains normative sections describing the use of the UN/EDIFACT syntax in SDMX messages.
3. SDMX-ML - the XML format for the exchange of SDMX-structured data and metadata. This document has normative sections describing the use of the XML syntax in SDMX messages, and is accompanied by a set of XML schemas and sample XML document instances.
4. The SDMX Format Implementor's Guide – this is a guide to help those who wish to use the SDMX format specifications. It includes reference material for the use of the SDMX Information Model; a section describing the expressive differences of the various messages and syntaxes; and provides some best practices for assigning identifiers and designing key families. This document is not normative.
5. Web Services Guidelines – this is a guide for those who wish to implement SDMX using web-services technologies. It places an emphasis on those aspects of web-services technologies which will work regardless of the development environment or platform used to create the web services, and recommends the use of the WS-I version 1.0 specification.

II. PROCESSES AND BUSINESS SCOPE

A. Process Patterns

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

1. *Bilateral exchange:* All aspects of the exchange process are agreed between 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.

2. *Gateway exchange:* Gateway exchanges are an organized set of bilateral exchanges, in which several data and metadata sending organizations or individuals agree to exchange the collected information with each other in a single, known format, and according to a single, known process. This pattern has the effect of reducing the burden of 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.
3. *Data-sharing exchange:* Open, freely available data formats and process patterns are known and standard. Thus, any organization or individual can use any counterparty's data and metadata (assuming they are permitted access to it). This model requires no bilateral agreement, but only requires that data and metadata providers and consumers adhere to the standards.

This document specifies the SDMX standards designed to facilitate exchanges based on any of these process patterns, and shows how SDMX offers advantages in all cases. It is possible to agree bilaterally to use a standard format (such as SDMX-EDI or SDMX-ML); it is possible for data senders in a gateway process to use a standard format for data exchange with each other, or with any data providers who agree to do so; it is certainly possible to agree to use the SDMX standards to support a common data-sharing process of exchange.

The standards specified here do not by themselves support a data-sharing process, but they provide the basic components of such a process. In the future, a more complete set of standards concerning the technology infrastructure for supporting a data-sharing process is expected be offered.

It is important to note that SDMX is primarily focused on the exchange of statistical data and metadata between providers and consumers. There may also be many uses for the standard model and formats specified here in the context of internal processing of data that are not concerned with the exchange between organizations and users. It is felt that a clear, standard formatting of data and metadata for the purposes of exchange can support any number of specific internal formats to facilitate processing by organizations and users.

B. 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 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 can be understood as the named dimensions of a multi-dimensional table, 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 metadata.

The SDMX standards offer a common model and formats which support the exchange of any type of statistical data meeting the definition above; an attempt has been made to optimize formats based on the specific requirements of each implementation, as described below in the SDMX-ML section.

The term "metadata" is very broad indeed, and the current version of SDMX is not designed to model or format every type of metadata. A distinction can be made between “structural” metadata – those concepts used in the description, identification and retrieval of statistical data – and “reference” metadata – that describe statistical concepts, methodologies for the generation of data and information on data quality.

Reference metadata, sometimes generated, collected or disseminated separately from the data to which they refer can be relevant to all instances of data described: entire collections of data, data sets from a given country, or for a data item concerning one country and one year. For this reason, some overlap may exist

between “reference” metadata – which are often disseminated separately from the data they refer – and “structural” metadata used to identify data.

The present SDMX specifications are concerned with the structural metadata needed to identify, use, and process the data "cubes" described above. Future releases of the SDMX specifications are envisaged to address the exchange of reference metadata in more detail, as the general scope of the SDMX-ML format is to enable the exchange of structured data and metadata needed by users. In this version 1.0, the term “metadata” is used narrowly, to describe what can be termed "structural" metadata.

It should be noted that these specifications are not intended to cover all possible views of data “cubes”. A “cube” is a rich, multi-dimensional construct, which can be viewed along any of its axis (or “dimensions”). SDMX takes a slightly narrower view of these requirements in its version 1.0 specifications. The view of data in these SDMX formats is primarily as time series – that is, as a set of observations which are organized around the time dimension, so that each observation occurs progressively through time. This is a clear way of organizing statistical data of many types, and has been proven to be a useful way of organizing data for exchange between counterparties.

There are, however, some types of statistical data which are not typically organized for exchange in this way – what we term “cross-sectional” data. SDMX provides some support for cross-sectional views of data cubes. In the 1.0 version of these standards, it is assumed that most data will be structured as time series. It is possible to describe a view such that a dimension other than time is used as the organizing dimension. If this type of a data structure is described – and if it has time as a dimension within the cube - then that data will be expressible in SDMX formats which are either organized along the chosen non-time dimension, or along time as a dimension. This approach gives time-series-based systems the ability to process many cross-sectional data sets as well as time series. In future versions of these specifications, it is intended that more flexible support for data cubes will be provided.

Another type of structure commonly found in statistical "cubes" of data is the use of hierarchical classifications to describe the points along any of its dimensions (or axes). In the 1.0 version, SDMX standards do not provide full support for this functionality. The introduction of these hierarchical classifications is anticipated in future versions of the standards.

C. 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:

1. *Batch Exchange of Data and Metadata:* The transmission of whole or partial databases between counterparties, including incremental updating.
2. *Provision of Easily Processible 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.
3. *Generic Processes:* While many applications and processes are specific to some 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.
4. *Presentation and Transformation of Data:* In order to make data and metadata useful to consumers, they must 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 must 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.