SDMX STANDARDS: SECTION 1

FRAMEWORK FOR
SDMX TECHNICAL STANDARDS

Version 3.0
DRAFT

May 2021

Revision History

Revision	Date	Contents
DRAFT 1.0	May 2021	Draft release updated for SDMX 3.0 for public consultation

Contents

1 Introduction

The Statistical Data and Metadata Exchange (SDMX) initiative (https://www.sdmx.org) sets standards that can facilitate the exchange of statistical data and metadata using modern information technology.

The SDMX Technical Specifications are organised into several discrete sections.

The following are published on the SDMX website (https://www.sdmx.org):

Section 1 **Framework for SDMX Technical Standards** – this document providing an introduction to the technical standards.

Section 2 **SDMX Information Model** - the SDMX information model is a standardised object model for modelling statistical domains centring on the structure of their data and metadata sets, the coding schemes used for classification, and the rules for controlling the exchange of data and metadata between organisations. This document provides a UML specification with supporting narrative.

Section 5 **SDMX Registry Specification** – an SDMX 'registry' acts as a repository for structural metadata and provisioning information, and a registry of data and metadata sources. This document sets out the specification.

Section 6 **SDMX Technical Notes** – detailed technical guidance for implementors of the SDMX standard.

The following are published on the GitHub repository of the SDMX Standards Technical Working Group (https://github.com/sdmx-twg):

sdmx-twg/sdmx-rest – REST API
Technical specifications for the SDMX RESTful web services application programming interfaces (API).

sdmx-twg/sdmx-ml – SDMX-ML
Technical specifications for the XML transmission format including XSD schemas, documentation and samples for data, structure and reference metadata messages.

sdmx-twg/sdmx-json – SDMX-JSON
Technical specifications for the JSON transmission format including documentation, schemas and samples for data, structure and reference metadata messages.

sdmx-twg/sdmx-csv – SDMX-CSV
Technical specifications for the SDMX-CSV transmission format for 'comma-separated values' (CSV) data and reference metadata.

The following sections are obsolete:
Section 3 - SDMX-ML - replaced by the sdmx-twg/sdmx-ml GitHub repository
Section 4 - SDMX-EDI
Section 7 - API - replaced by the sdmx-twg/sdmx-rest GitHub repository

VTL

In July 2020 the SDMX 2.1 specifications were revised to add support for the Validation and Transformation Language (VTL). For 3.0, the VTL specification has been updated to align with changes to the information model and other modifications to the Standard such as the introduction of Semantic Versioning for the versioning of structural metadata artefacts. Section 2 (Information Model) sets out details of the 'Transformation and Expressions' package for defining and managing VTL 2.0 programs and Section 6 (Technical Notes) provides detailed guidance on implementing and using VTL with SDMX.

2 Change History

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

The 2.1 version of this standard represents a set of changes resulting from several years of implementation experience with the 2.0 standard. 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 3.0 version incorporates new features, improvements and changes arising from the collective knowledge gained from a decade of operating experience with the 2.1 standard. In pursuit of modernisation and simplification, features considered obsolete have been deprecated – in particular the EDI transmission format, the lesser-used XML data messages and the SOAP web services API. Many areas remain backwardly compatible with 2.1, but there are some breaking changes where the information model has been redesigned to better support practical use case. Structure mapping and reference metadata are examples. The opportunity has been taken to revise the RESTful web services API which is also not backwardly compatible, but benefits from a rationalisation and better organisations of resources, and a much richer data query URL syntax.

2.1 Major Changes from 1.0 to 2.0

- **Reference Metadata**: In addition to describing and specifying data structures and formats (along with related structural metadata), the version 2.0 specification also provides for the exchange of metadata which is distinct from the structural metadata in the 1.0 version. 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 the 2.0 standard 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.

2.2 Major Changes from 2.0 to 2.1

- **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. These changes are largely to leverage the Cross-domain Concepts of the Content Oriented Guidelines.
- **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 in the registry interfaces and elsewhere. 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 (version 2.0 Compact) and non-time series data sets (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**.**

2.3 Major Changes from 2.1 to 3.0

SDMX version 3.0 introduces new features, improvements and changes to the Standard in the following key areas:

Information Model

- Simplification and improvement of the reference metadata model
- Support for microdata
- Support for geospatial data
- Support for code list extension and discriminated union of code lists
- Improvements to structure mapping
- Improvements to code hierarchies for data discovery
- Improvements to constraints

Versioning of Structural Metadata Artefacts

- Adoption of the three-number semantic versioning standard for structural metadata artefacts (https://semver.org)

REST Web Services Application Programming Interface (API)

- Change to a single 'structure' resource for structure queries simplifying the REST API specification by reducing the number of resources to five
- Improvements to data queries
- Improvements to reference metadata queries
- Support for structural metadata maintenance using HTTP PUT, POST and DELETE verbs

SOAP Web Services API

- The SOAP web services API has been deprecated with version 3.0 standardising on REST

XML, JSON, CSV and EDI Transmission formats

- The SDMX-ML, SDMX-JSON and SDMX-CSV specifications have been extended and modified where needed to support the new features and changes such as reference metadata and microdata
- The SDMX-EDI transmission format for structures and data has been deprecated
- Obsolete SDMX-ML data messages including Generic, Generic Time Series and Structure Specific Time Series have been deprecated standardising on Structure Specific Data as the sole XML format for data exchange

SDMX 3.0 messages may continue to be converted between transmission formats without loss of information.

Several of the changes are ‘breaking’ meaning that, in specific cases, the version 3.0 specification is not backwardly compatible with earlier versions of the Standard.

The SDMX 3.0 Major Changes document provides more information including an analysis of the breaking changes.

3 Processes and Business Scope

3.1 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 collecting 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-ML or SDMX-JSON); 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 possible to agree to use the full set of SDMX standards to support a common data-sharing process of exchange, whether based on an SDMX-conformant registry or some other architecture.

The standards specified here specifically support a data-sharing process 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 processing. 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 is 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.

It should be pointed out that these different process models are not mutually exclusive – a single system capable of expressing data and metadata in SDMX-conformant formats could support all three scenarios. Different standards may be applicable to different processes (for example, many registry services interfaces are used only in a data-sharing scenario) but all have a common basis in a shared information model.

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. But 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 – it becomes more easily possible to process a standard format once the data is obtained, and the data and metadata are linked together, making the comprehension and 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.

It is important to note that SDMX is primarily focused on the *exchange* and *dissemination* of statistical data and metadata. 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, however. It is felt that a clear, standard formatting of data and metadata for the purposes of exchange and dissemination can also facilitate internal processing by organizations and users, but this is not the focus of the specification.

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:

1. *Batch Exchange of Data and Metadata:* The transmission of whole or partial databases between counterparties, including incremental updating.

2. *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, JSON and REST 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.

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 co-ordinate 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 “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 is 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” – metadata which are useful to those who need to understand the content and form of a data provider’s output. 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 classification schemes in organizing and managing the exchange and dissemination of data and metadata. It is possible to express information about classification schemes and domain categories in SDMX, along with their relationships to data and metadata sets, as well as to categorize other objects in the model.

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 schematically in Figure 1, and are discussed in more detail elsewhere in this document.

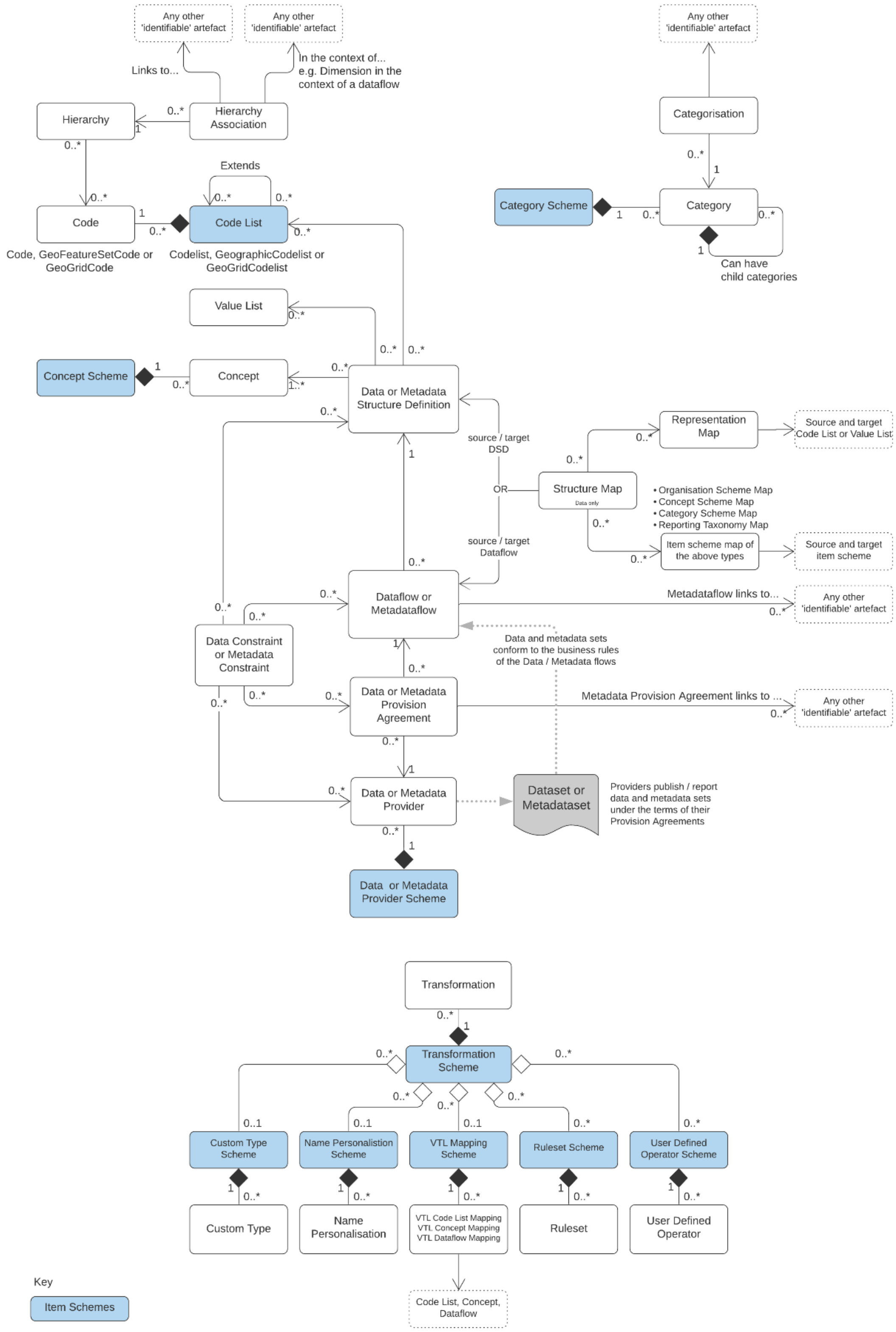

Figure 1: High Level Schematic of Major Artefacts in the SDMX 3.0 Information Model

3.4 The SDMX View of Statistical Exchange

Version 1.0 of ISO/TS 17369 SDMX covered statistical data sets and the metadata related to the structure of these data sets. This scope was useful in supporting the different models of statistical exchange (bilateral exchange, gateway exchange, and data-sharing) but was not by itself sufficient to support them completely. Versions 2.0 and 2.1 provide a much more complete view of statistical exchange, so that an open data-sharing model can be fully supported, and other models 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 the set of formal objects not only in a technical sense, but also in terms of what they represent in the real-world exchange of statistical data and metadata.

The first version of SDMX 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 are retained and enhanced in the second version of the standards 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. Data structures are owned and maintained by agencies - in a similar fashion, data flows are owned by maintenance agencies.

SDMX allows for the publication of statistical data (and the related structural metadata) but also provided for the standard, systematic representation of reference metadata. In version 2.1, reference metadata were reported independent of the statistical data. However, in 3.0 reference metadata associated directly with data such as footnotes are reported as attributes of the data set. For other reference metadata, principally that linked to structures like “concepts”, SDMX provides reference "metadata sets", "metadata structure definitions", and "metadata flows". These objects are very similar to data sets, data structure definitions, and data flows, but 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" which is maintained by a specific agency. Concept schemes group a set of concepts, provide their definitions and names, and allow 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 - key families - and for reference metadata structures.

Inherent in any statistical exchange – and in many dissemination activities – is a concept of "service level agreement", even if this is not formalized or made explicit. SDMX incorporates this idea in objects termed "provision agreements". 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 which tell you which data providers are supplying what data to which data flows. Similarly, metadata provision agreements for metadata flows.

Provision agreements allow for a variety of information to be made available: the schedule by which statistical data or metadata is reported or published, the specific topics about which data or metadata is reported within the theoretically possible set of data (as described by a data structure definition or reference metadata structure definition), and 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:

- ***Data Set:*** Data is 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 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 Version 2.1, the term "Key Family" was replaced by "Data Structure Definition" (DSD) both in XML Schemas and the Information Model. The DSD has been modified in version 3.0 to better support microdata by providing the option to define multiple measures and for attributes and measures to take arrays of values.
- ***Code list:*** Code lists enumerate a set of codes to be used in the representation of dimensions, attributes, and other structural parts of SDMX. Codes can be organised into simple hierarchies within a code list, and more complex hierarchies potentially involving multiple code lists using hierarchy and hierarchy association structures.
- ***Value list:*** Value lists introduced in version 3.0 are similar to codelists with the exception that the items do not need to conform to the usual SDMX rules for identifiable objects. That allows the values to include characters such as currency symbols (e.g. ¥) which would otherwise make illegal codes. However, unlike codes, values are not individually identifiable. Value lists find application in concepts and data structures definitions for less structured data and microdata enumerations and can be mapped to other value or code lists using representation maps.
- ***Organisation Scheme:*** Organisations and organisation structure can be defined in an Organisation Scheme. Specific Organisation Schemes exist for Maintenance Agency, Data Provider, Metadata Provider, Data Consumer, and Organisation Unit.
- ***Category Scheme and Categorisation:*** 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 Categorisation links a category to an identifiable object. In this way sets of objects can be categorised. 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"). Note that this core

representation can be overridden in the data structure definition or metadata structure definition that uses the concept. Indeed, organisations wishing to remain with version 1.0 key family schema specifications will continue to declare the representation in the key family definition.

- ***Metadata Set:*** A reference metadata set is a set of information pertaining to an object within the formal SDMX view of statistical exchange: they may describe the maintainers of data or structural definitions; they may describe the schedule on which data is released; they may describe the flow of a single type of data over time; they 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), what their presentational structure is, 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 Categorisation (this facilitates the search for data according to organised 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, categorises, and constrains metadata sets.
- ***Data Provider:*** An organization which produces data is termed a data provider.
- ***Metadata Provider:*** An organization which produces reference metadata is termed a metadata provider.
- ***Provision Agreement (Metadata Provision Agreement):*** The set of information which describes the way in which data sets and metadata sets are provided by a data/metadata 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 (which is outside the scope of a technical specification). In version 3.0, metadata provision agreement and data provision agreement are two separate artefacts.
- ***Constraint:*** Data and Metadata Constraints describe a subset of a data source or metadata source, and may also provide information about scheduled releases of data. They are associated with data / metadata providers, provision agreements, data flows, metadataflows, data structure definitions and metadata structure definitions.
- ***Structure Map:*** Structure maps describes a mapping between data structure definitions or dataflows for the purpose of transforming a data set into a different structure. The mapping rules are defined using one or more component maps which each map in turn describes how one or more components from the source data structure definition map to one or more components in that of the target. Represent

maps act as lookup tables and specific provision is made for mapping dates and times.

- ***Representation Map:*** Representation maps describe mappings between source value(s) and target value(s) where the values are restricted to those in a code list, value list or be of a certain type such as integer or string.
- ***Item Scheme Map:*** An item scheme map describes mapping rules between any item scheme with the exception of code lists and value lists which use representation maps. The version 3.0 information model provides four item scheme maps: organisation scheme map, concept scheme map, category scheme map and reporting taxonomy map.
- ***Reporting Taxonomy:*** A reporting taxonomy allows an organisation 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:*** The process class 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.
- ***Hierarchy***: Describes complex code hierarchies principally for data discovery purposes. The codes themselves are referenced from the code lists in which they are maintained.
- ***Hierarchy Association***: A hierarchy association links a hierarchy to something that needs it like a dimension. Furthermore, the linking can be specified in the context of another object such as a dimension in the context of a dataflow. Thus, a dimension in a data structure definition could have different hierarchies depending on the dataflow.
- ***Transformation Scheme:*** A transformation scheme is a set of Validation and Transformation Language (VTL) transformations aimed at obtaining some meaningful results for the user (e.g., the validation of one or more data sets). The set of transformations is meant to be executed together (in the same run) and may contain any number of transformations in order to produce any number of results. Thus, a transformation scheme can be considered as a VTL 'program'.

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, it is felt that an architecture based on a set of registry services is potentially useful. A "registry" – as understood in webservices 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 (though note that the mechanism of access control is outside of the scope of the SDMX standard). It can be understood as the index of a distributed database or metadata repository which is made up of all the data provider's data sets and reference metadata sets within a statistical community, located across the Internet or similar network.

Note that the SDMX registry services are not concerned with the storage of data or reference metadata. The assumption is that data and reference metadata lives on the sites of its data and metadata 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 an SDMX REST API query message or other mechanism).

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, provided the interfaces are supported as specified in Section 5 on the Registry Specification. These interfaces are expressed as XML documents, but also REST API request/response messages

The registry services discussed here can be briefly summarized:

- ***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 sets and reference metadata sets. The registry stores metadata about these objects, 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. Objects in the registry are organized and categorized according to one or more category schemes.
- ***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 sets and reference metadata sets, 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 all subscribers whenever the registry objects are updated.

3.6 RESTful 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: for requesting and supplying data; for expressing the enveloping data which is used to package exchanged data; for describing web services to one another, to allow for easy integration into applications that use other web services as data resources.

Version 3.0 has standardized on RESTful web services with a OpenAPI specification published on the SDMX Technical Working Group's GitHub repository https://github.com/sdmx-twg. There are five 'resources':

- structure – retrieval and maintenance of structural metadata
- data – retrieval of data
- schema – retrieval of XML schemas to validate specific data or metadata sets
- availability – retrieval of information on the data available for a Dataflow
- metadata – retrieval of reference metadata

The following conceptual example uses the 'data' resource to query a data repository for a series identified by the key 'M.USD.EUR.SP00.A' in the EXR (ECB exchange rates) Dataflow:

```
https://ws-entry-point/data/dataflow/ECB/EXR/1.0.0/M.USD.EUR.SP00.A
```

4 The SDMX Information Model

SDMX provides a way of modelling statistical data, and defines the set of metadata constructs used for this purpose. Because SDMX specifies a number of transmission formats for expressing data and structural metadata, the model is used as a mechanism for guaranteeing that transformation between the different formats is lossless. In this sense, all of the formats are syntax-bound expressions of the common information model.

SDMX recognizes that statistical data is structured; in SDMX this structure is termed a Data Structure Definition. "Data sets" are made up of one or more lower-level "groups", based on their degrees of similarity. Each group is in turn comprised of one or more "series" of data. Each series or section has a "key" - 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.

There is some similarity between “cube” structures commonly used to process statistical data, and the Data Structure Definition idea in the SDMX Information Model. It is important to note that the data as structured according to the SDMX Information Model is optimized for exchange, potentially with partners who may have no ability to process a “cube” of data coming from complex statistical systems. SDMX time series can be understood as “slices” of the cube. Such a slice is identified by its key. A "series" key consists of the values for all dimensions specified by the key family except time. Thus, it is possible to reconstruct and describe data cubes from SDMX-structured data, and to exchange such databases using the interfaces and formats provided for that purpose in the standard. Additional objects such as hierarchical code lists, constraints and structure maps make it possible to more fully model the structure of cubes.

The information model also provides a view of reference metadata: a mechanism for referencing the meaningful “objects” within the SDMX view of statistical exchange processes (data providers, structures, provisioning agreements, dataflows, metadata flows, etc.) to which metadata is attached; a mechanism for describing a set of meaningful concepts, of organizing them into a presentational structure, and of indicating how their values are represented. This is based on a simple, hierarchical view of reference metadata which is common to many metadata systems and classification/categorization schemes. SDMX provides a model (and XML and JSON formats) for both describing reference metadata structures, and of reporting reference metadata according to those structures.

Version 2.0/2.1 introduced support for metadata related to the process aspects of statistical exchange. A step-by-step process can be modelled; information about who is providing data and reference metadata and how they are providing it can be expressed; and the technical aspects of service-level agreements (and similar types of provisioning agreements) can be represented.

Support for the Validation and Transformation Language (VTL) in the SDMX Information Model was introduced in the July 2020 revision of 2.1 and is retained in version 3.0 with minimal changes. This allows reusable VTL ‘programs’ (a cohesive set of transformation statements designed to be executed together) and their associated constructs such as validation rulesets and user-defined operators to be managed and exchanged as SDMX structural metadata. Mappings between objects such as data sets referenced in VTL programs and the actual SDMX artefacts to which they relate is essential when it comes to actually executing programs, and this information can also be defined. Chapter 7 has more information on VTL and its integration with SDMX.

A full UML conceptual design of the information model is set out in Section 2 of the Technical Specifications.

5 The SDMX Transmission Formats

5.1 SDMX-ML

SDMX-ML is the XML transmission format specification for exchanging structural metadata, data and reference metadata, and interacting with SDMX registry services. It is designed as a general-purpose format for all automation and data / metadata exchange tasks, and provides the most complete coverage.

There are four distinct types of message:

1. *Structure Definition:* For the exchange of structural metadata. A SDMX-ML structure message can carry details of any number and combination of structural metadata artefacts like DSDs, code lists and constraints.

2. *Structure-specific Data:* For the exchange of data. This format is specific to the Data Structure Definitions of the data sets (in other terms, it is DSD-specific) and is created by following mappings between the metadata constructs defined in the Structure Definition message and the technical specification of the format. It supports the exchange of large data sets in XML format, 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.

   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.

3. *Generic Metadata:* For the exchange of reference metadata sets. 'Generic' means the XML elements and XML attributes are the same regardless of the metadata set.

4. *Registry:* All of the possible interactions with the SDMX registry services are supported using SDMX-ML interfaces and REST API calls. Submission of structural metadata content, data / metadata registrations and subscriptions is performed by a synchronous exchange of documents – a "request" message answered by a "response" message.

5.2 SDMX-JSON

SDMX-JSON is the JSON transmission format specification for exchanging structural metadata, data and reference metadata. It provides an alternative to SDMX-ML and is most suited to applications like web data dissemination.

SDMX-JSON messages serve the same function as those of the XML formats but have a different structure. For data, an important distinction is that they carry both component codes and labels which provides all the information needed to display the content in a single JSON response. The XML Structure-specific Data format by contrast carries only code IDs thus requiring applications obtain and hold structural metadata about the data set in order to display the content in human-readable form.

SDMX-JSON does not include messages for subscription / notification or registration registry services - SDMX-ML must be used for those purposes.

There are three distinct message types:

1. *Structure:* For the exchange structural metadata. SDMX-JSON structure messages follow the same principles as for SDMX-ML in that a single message can transmit any

number and combination of structural metadata artefacts. While the SDMX-ML and SDMX-JSON messages are structured differently, it is possible to freely convert between them.

2. *Data:* For the exchange of data. Unlike SDMX-ML, the structure of a SDMX-JSON data message is not specific to the DSDs of the data sets so schema validation will not check for compliance of the data with the DSDs.

3. *Metadata*: For the exchange of reference metadata sets.

5.3 SDMX-CSV

SDMX-CSV is the CSV transmission format specification for exchanging data and reference metadata only.

SDMX-CSV provides a simple columnar format for data and metadata that can be readily created and interpreted by standard software tools such as Microsoft Excel. Nevertheless, data and metadata can still be converted between the CSV and the JSON / XML formats without loss.

There are two distinct message types:

1. *Data*: For the exchange of data. Like SDMX-JSON, SDMX-CSV can include both code IDs and labels which is helpful when using the data to create human readable charts and dashboards.

2. *Metadata*: For the exchange of reference metadata sets.

5.4 Formats and Messages Deprecated in Version 3.0

The following formats and messages have been deprecated in version 3.0 to simplify, modernise and rationalise the standard.

- SDMX-EDI
- SDMX-ML 1.0/2.0 Generic (time-series) data message
- SDMX-ML 1.0/2.0 Compact (time-series) data message
- SDMX-ML 1.0/2.0 Utility (time-series) data message
- SDMX-ML 1.0/2.0 Cross-Sectional data message
- SDMX-ML 2.1 Generic (Time Series) data messages (for observations, time-series and cross-sectional data)
- SDMX-ML 2.1 Structure Specific Time Series data message

The following messages were deprecated in version 3.0 as a consequence of the deprecation of the SOAP web services:

- SDMX-ML Query messages
- SDMX-ML Submit Structure Request messages

6 Dependencies on SDMX content-oriented guidelines

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. There are three major parts to this effort.

6.1 Cross-Domain Concepts

The SDMX Cross-Domain Concepts is a content guideline concerning concepts which are used across statistical domains. This list 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 organisations 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 two-fold: to provide a core set of concepts which can be used to structure statistical data and metadata, to promote interoperability between systems ("structural metadata", as described above); and 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.)

6.2 Metadata Common Vocabulary

The Metadata Common Vocabulary 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 ISO-compliant 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.

6.3 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, 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 organisation of the content of such schemes is supported in SDMX as a Category Scheme.

SDMX Statistical Subject-Matter Domains will be listed and maintained by the SDMX Initiative and will be subject to adjustment.

6.4 SDMX Concept Roles

These guidelines define the standard set of SDMX Concept Roles and their use. This set of standard SDMX Concepts are implemented as a cross-domain Concept Scheme that defines the set of concept roles and gives examples on concept role implementation in SDMX 2.0, 2.1 and 3.0. A concept role gives a particular context to a concept for easy and systematic interpretation by machine processing and visualization tools. For example, the concepts REPORTING_AREA and COUNTERPART_AREA are different concepts but they are both geographical characteristics, therefore they can be associated with the same concept role ID: "GEO". This allows visualization systems to interpret these concepts as geographical data in order to generate maps. The implementation of concept roles is different in versions 2.0 and 2.1/3.0 of the SDMX technical standard. Specifically for SDMX 3.0, this set of roles is considered a normative list that must be interpreted in the same way by all organisations. Additional roles may be provided via the standard roles' mechanism in SDMX 3.0, i.e., via Concept Schemes; the semantics of these roles have to be agreed bilateraly in data exchanges. The Concept Roles are available as an SDMX Concept Scheme on the SDMX Global Registry.

7 Validation and Transformation Language

For many years the SDMX initiative has been fostering and supporting the development of a standard calculation language, called Validation and Transformation Language (VTL). A blueprint for defining calculations was already described in the original SDMX 2.1 specifications (package 13 of the Information Model - "Transformations and Expressions"). It was just a basic framework that required further developments to became operational in order to achieve a calculation language able to manipulate SDMX artefacts.

These developments started in late 2012 and were put in charge of the Validation and Transformation Language Task Force (VTL TF), which included members of the SDMX Technical Working Group (TWG) and Statistical Working Group (SWG), besides experts coming from the DDI and GSIM communities. The intent was to define a standard language to be implemented in SDMX and applicable also to GSIM and DDI. This brought to the publication of the VTL 1.0 in 2015. Then new requirements came from a number of proofs of concepts and tests of VTL 1.0 made by several organisations and triggered a large improvement of the language. A new provisional version, the VTL 1.1, was released in public consultation in 2017. The high number of comments received triggered another phase of intensive work, with the main goal of achieving a more robust and forward compatible version. Finally, the VTL 2.0 was published between April and July 2018 (see the SDMX website).

The implementation of the VTL 2.0 in SDMX started in late 2018 and was published as an incremental revision to the SDMX 2.1 standards in July 2020. It allows users to write VTL 2.0 programs for validating and transforming SDMX data, to store these programs in a SDMX metadata registry and to exchange them through SDMX messages, also together the definition of the data structures of the involved data.

The Transformations and Expressions package for modelling VTL programs in the SDMX information model is explained in Section 2 of the Technical Specifications with further detailed usage and implementation guidance given in Section 6.