SDMX self-learning package No. 5

Student book

Metadata Structure Definition

Produced by	Eurostat, Directorate B: Statistical Methodologies and Tools Unit B-5: Statistical Information Technologies
Last update of content	December 2010
Version	1.0

TABLE OF CONTENTS

1 Scope of the Student Book

1.1 Purpose

This student book will provide a complete understanding of how metadata structure definitions relate to the SDMX standard; including their relation to the information model, reference metadata, and the parallels to data structure definitions. By the end of this book, the user should be able to understand the role of metadata structure definitions and how to create them.

1.2 Context

This student book is the fifth in a set of student books (see Table 1 - Student books on SDMX), which together provide a complete understanding of SDMX.

Ref.	Title
[01]	Introduction to SDMX
[02]	The SDMX Information Model
[03]	SDMX-ML Messages
[04]	Data Structure Definition
[05]	Metadata Structure Definitions
[06]	XML based technologies used in SDMX
[07]	SDMX architecture using the pull method for data sharing – Part 1
[08]	SDMX architecture using the pull method for data sharing – Part 2

Table 1 - Student books on SDMX

1.3 Prerequisites

The reading of the first and second student books are necessary in order to gain a basic understanding of the SDMX standard and its purpose, as well as the fundamental principles of the SDXM information model which will be built upon here. In addition, it is strongly recommended that the fourth book be read in order to understand the parallels between the data and metadata structure definitions.

2 Background Information

2.1 Scope of the Chapter

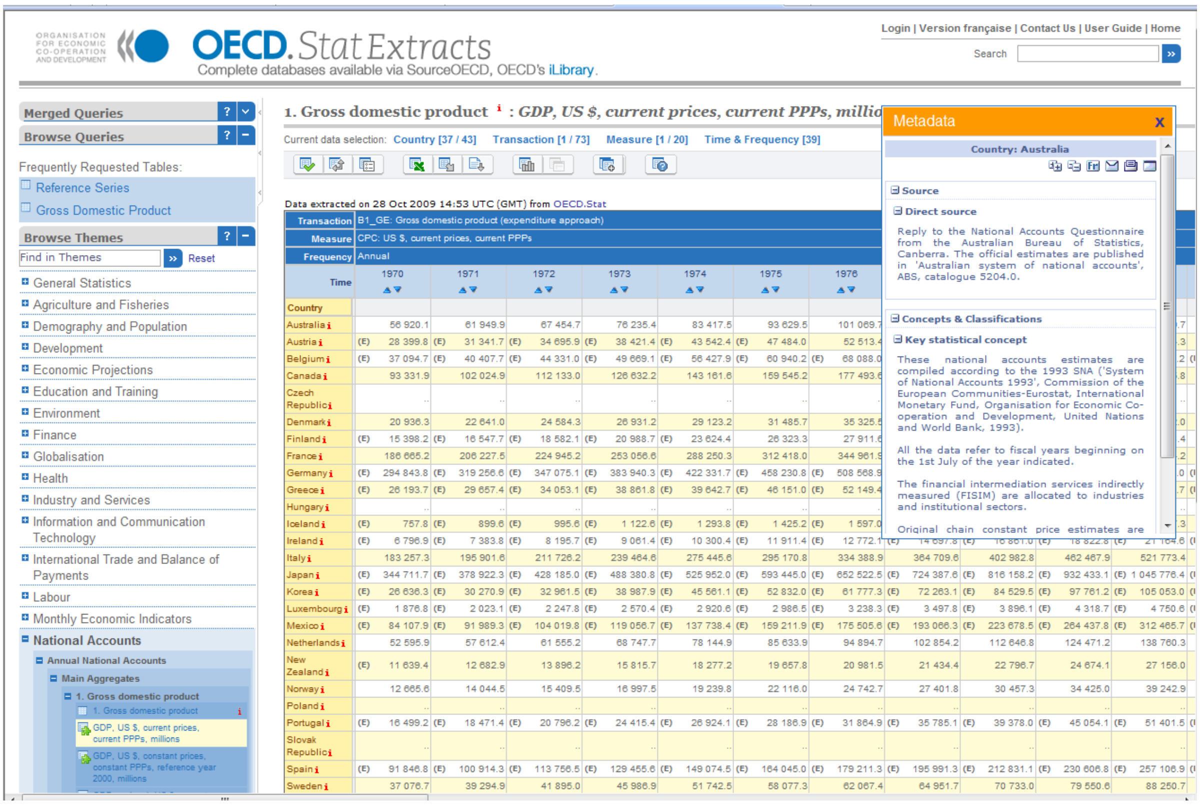
This chapter provides high level background information about reference metadata, and specifically about quality frameworks. It discusses how the reference metadata is used, how the metadata structure definition relates to it, and how it fits into the SDMX Information Model. This is not intended to be exhaustive discussion of the information model or quality frameworks: the intention is to merely to provide enough information to understand the purpose SDMX metadata structure definitions serve in supporting metadata systems..

2.2 Purpose of Reference Metadata

In the SDMX metamodel, objects can have descriptive metadata associated with them. This is typically, and for the purposes of this book, described as reference metadata. The SDMX standard allows for reference metadata to be stored and exchanged without it being embedded in the object it is describing. In other words, such metadata is linked to the object by a reference to the object. Further SDMX dictates that this metadata should be indexable for searching, and structured for ease of processing and reporting. To achieve this, the information model contains a metadata set and a metadata structure definition. The metadata set contains the metadata and a reference to the object to which the metadata pertains. The metadata structure definition specifies how the reference metadata is to be indexed and structured for processing and reporting.

As the reference metadata comprises SDMX attributes (called metadata attribute in the model) and a data set can also contain attributes (called data attributes in the information model) there is an issue of how can one decide between what can be contained in a data attribute in a Data Structure Definition (used to support the reporting, exchange and dissemination of data), and what should be contained in a metadata attribute in a Metadata Structure Definition (used to support the reporting, exchange and dissemination of reference metadata).

The guiding rule here is to analyse how the metadata are collected and used. Often, reference metadata relate not to a single set of data but to a whole data collection and dissemination process, perhaps broken down by subject domain and organisation. Such metadata are not appropriate to be reported or exchanged with the data, as that do not relate directly to the data. These sorts of data are often referred to as Quality Framework metadata such as the IMF SDDS and GDDS, Eurostat ESMS, and the OECD Quality Framework. An example of how such metadata may be used in a data dissemination system is shown in the screen shot below. Here, the metadata are about a particular country’s statistical system for the chosen statistical domain, and not about any specific set of data.



Source: OECD

Figure 1: Screen shot showing metadata linked to the data collection process

On the other hand consider the unit of measure and the unit multiplier. Without this information, one cannot effectively process the observations contained in data set - since very little can be actually understood about the values. Information such as this should be contained in data attributes and should accompany the data directly in a data set.

Another consideration is the nature of the information. Data attributes are singular in nature and there is no grouping mechanism in a Data Structure Definition. Therefore, it would not be practical to exchange, for instance, contact information within a data set, except as a single text string. In the example that will be presented in the next section, contact information is supplied and there it is done by specifying individual components of the contact details (such as name, phone number, address, etc.). Reference metadata allows such detail to be captured.

2.3 Reference Metadata Example

For the purpose of this book an example of reference metadata has been taken from the Eurostat website, in the national accounts main statistics tables. These tables can be found at the URL:
http://epp.eurostat.ec.europa.eu/portal/page/portal/national_accounts/data/main_tables.

For convenience, a capture of the relevant information is provided in Figure 1. First, it is important to understand the manner in which the data is structured. The data sets are organised into categories. As you can see in this navigation tree, certain nodes have an icon ( ) next to them. These icons indicate that there is metadata associated the data set or collection of data sets for a particular category.

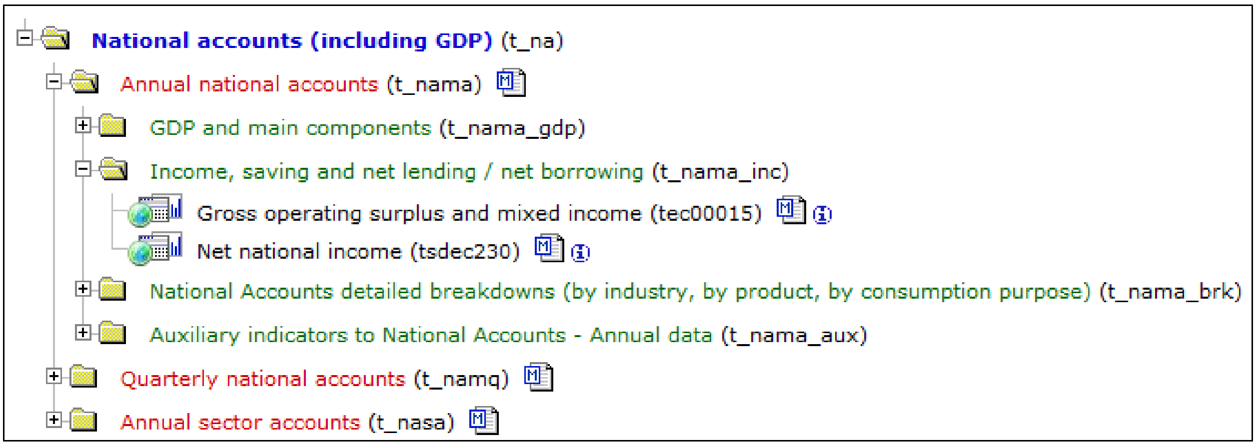


Figure 2: Eurostat national accounts statistical tables navigation tree

Clicking on one of these icons at a directory level will open a new window containing information (see in Figure 3). This information is reference metadata pertaining to the category. In this context, this information applies to all the data sets for this category. It can be seen that the reference metadata contains descriptive information such as contact details, information about when the reference metadata was last updated, and statistical presentation information.

The figure below is an extract from the Euro-SDMX Metadata Structure (ESMS) list of Concepts. These terms are used in the examples in this book.

EURO-SDMX Metadata Structure (release 3, March 2009)

	Concept Name	Concept Code	Descriptions	Representation	ESS Guidelines
1	**Contact**	CONTACT	Individual or organisational contact points for the data or metadata, including information on how to reach the contact points.		
1.1	Contact organisation	CONTACT_ORGANISATION	The name of the organisation of the contact points for the data or metadata.	Text	The full name of your organisation.
1.2	Contact organisation unit	ORGANISATION_UNIT	An addressable subdivision of an organisation.	Text	The name of the unit responsible for the metadata file (it can also include a unit number).
1.3	Contact name	CONTACT_NAME	The name of the contact points for the data or metadata.	Text	The name of the person responsible for the statistical domain (first name and family name).
1.4	Contact person function	CONTACT_FUNCT	The area of technical responsibility of the contact, such as "methodology", "database management" or "dissemination".	Text	The title of the person responsible for the statistical domain (this title can contain the precise area of responsibility such as methodologist or data base manager)
1.5	Contact mail address	CONTACT_MAIL	The postal address of the contact points for the data or metadata.	Text	The postal address of the person responsible for the statistical domain.
1.6	Contact email address	CONTACT_EMAIL	E-mail address of the contact points for the data or metadata.	e-mail	The email address of the person responsible for the statistical domain (this can be an individual mail address or a functional mailbox).
1.7	Contact phone number	CONTACT_PHONE	The telephone number of the contact points for the data or metadata.	Telephone	The phone number of the person responsible for the statistical domain.
1.8	Contact fax number	CONTACT_FAX	Fax number of the contact points for the data or metadata.	Fax	The fax number of the person responsible for the statistical domain.
2	**Metadata update**	META_UPDATE	The date on which the metadata element was inserted or modified in the database.		
2.1	Metadata last certified	META_CERTIFIED	Date of the latest certification provided by the domain manager to confirm that the metadata posted are still up-to-date, even if the content has not been amended.	Date	The date of the latest certification of this metadata file in order to confirm that the metadata file produced is still up-to-date. Such a certification can also be done if the contents of the metadata file has not been amended.
2.2	Metadata last posted	META_POSTED	Date of the latest dissemination of the metadata.	Date	The date when this metadata file is disseminated will normally be inserted automatically by the reference metadata production system (for Eurostat: by EMIS).
2.3	Metadata last update	META_LAST_UPDATE	Date of last update of the content of the metadata.	Date	The date when this metadata file is last updated will normally also be inserted by the reference metadata production system (for Eurostat: by EMIS).
3	**Statistical presentation**	STAT_PRES			
3.1	Data description	DATA_DESCR	Main characteristics of the data set described in an easily understandable manner, referring to the data and indicators disseminated.	Text	Describe the main characteristics of the data set in an easily understandable manner, referring to the main data and indicators disseminated. This short description should be understood immediately and easily by the users.
3.2	Classification system	CLASS_SYSTEM	Arrangement or division of objects into groups based on characteristics which the objects have in common	Text	List all classifications which are used for the data set produced (with their detailed names).

Figure 3: Extract of ESMS Concepts

Annual national accounts

Reference Metadata in Euro SDMX Metadata Structure (ESMS)

Compiling agency: Statistical Office of the European Communities (Eurostat)

For any question on data and metadata, please contact: EUROPEAN STATISTICAL DATA SUPPORT

1. Contact	
1.1 Contact organisation	Statistical Office of the European Communities (Eurostat)
1.2 Contact organisation unit	Unit C2: National accounts - production
1.5 Contact mail address	2920 Luxembourg LUXEMBOURG

2. Metadata update	
2.1 Metadata last certified	05 February 2009
2.2 Metadata last posted	05 February 2009
2.3 Metadata last update	05 February 2009

3. Statistical presentation
3.1 Data description

Figure 4: Extract of Eurostat annual national accounts reference metadata

This metadata is reported using the concepts of the ESMS and is structured according to the ESMS.

2.4 Purpose of Metadata Structure Definitions

Much like the relationship between data structure definitions and data sets, metadata structure definitions are structural metadata that describe the structure of metadata sets. A metadata structure definition defines the concepts and their hierarchy which comprise the metadata to be reported, the types of objects to which the metadata relate, and the means for identifying those objects. An effective metadata structure definition will allow metadata aware systems to understand the areas in which additional metadata may be available, and to create efficient systems for collecting and presenting the reference metadata.

In the example provided in the previous section, the icons are presented in the tree because the metadata structure definition allows the website to efficiently discover that there is reference metadata pertaining to the category. The reference metadata window can present the information in a clean and organised manner because the metadata structure definition defines the structure of the report and provides useful names and descriptions for the information being reported. Such metadata could be made available as a download (in much the same way as data are made available). In short, the ESMS example demonstrates the utility of metadata structure definitions in support of such quality frameworks.

2.5 Information Model

The metadata structure definition is best thought of as being analogous to a data structure definition. The information model intentionally models the metadata structure definition in a way that is similar to the data structure definition.

A data structure definition comprises dimensions (which, combined, comprise the key of the series for which observation or data attributes pertain), data attributes which are ‘attached’ to a part of the data set, and one or more measures which specify the observations. A metadata structure definition comprises identifier components (which, combined, comprise the key of the object to which the metadata pertains), and metadata attributes which are ‘attached’ to the object. Multiple ‘target objects’ can be specified in a metadata structure definition.

A data set (described by a data structure definition) contains keys and observations related to specific time periods, and data attributes that give additional information important to understanding and processing of the data. A metadata set (described by a metadata structure definition) comprises the key of an object to which metadata attributes are attached.

Like the data structure definition (or Key Family), the metadata structure definition inherits from the Structure base in the SDMX information model. This means that it is a maintainable object, which contains a collection of component lists known as groupings. In a metadata structure definition, these groupings are the Full Target Identifier, the partial target identifiers, and the report structures. In the data structure definition the groupings are the attribute list, dimension list, measure list, and groups. There are very strong correlations between the groupings of the data structure definition and the groupings of the metadata structure definition.

Where the dimension list (and to some extent groups) in the data structure definition contain dimensions which define how a data set describes what is being measured, the full and partial target identifiers in the metadata structure definition contain identifier components which define how a metadata set identifies what object is being described by the reference metadata.

In the data structure definition the attribute list and measure list contain attributes, measures, and attachment information which describes what information is in the data set and how it is presented. Similarly, the report structures in the metadata structure definition contain metadata attributes which describe what concepts are included in the reference metadata set. Further, the report structure defines how the metadata attributes are organised.

3 Step by Step Metadata Structure Definition Creation

3.1 Scope of Chapter

This chapter will work through an example of defining a metadata structure definition similar to the ESMS sample cited in Chapter 2. The example will be built up one grouping at a time. Before demonstrating the creation of a grouping, its general purpose will be described. Next, the components which make up the grouping will be described. After the grouping has been fully introduced, a demonstration of creating the grouping will be given, after which the resulting XML will be shown and described. Finally, a brief discussion of the known limitations of the current implementation will be discussed.

The demonstration will use the SDMX Metadata Editor, which is freely available from the Metadata Technology tools download site (http://www.metadatatechnology.com/software/). The examples assume that the basic use of this tool is understood. For details on how to perform specific actions within the tool, see the embedded help in the application.

3.2 Target Identifiers (Full and Partial)

3.2.1 Schematic

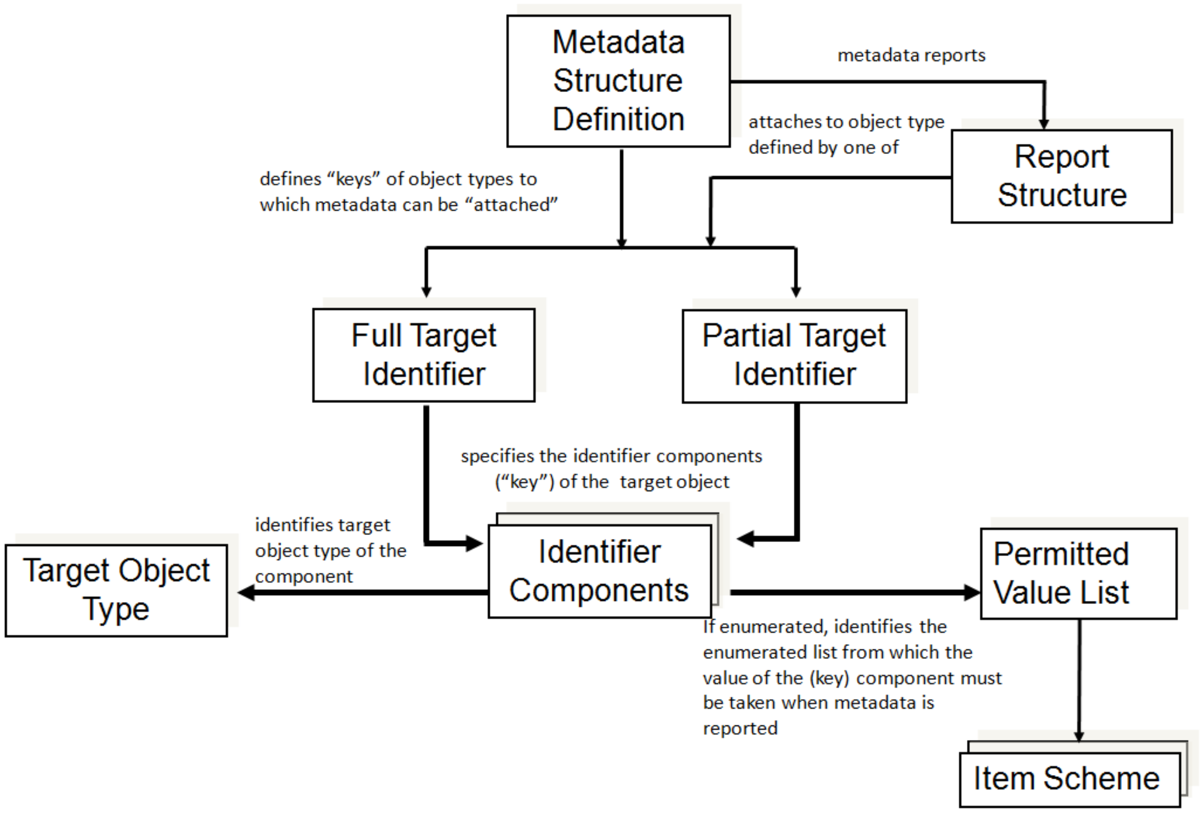


Figure 5: Schematic of the Information Model for Metadata Structure Definition

The metadata structure definition defines the target objects to which metadata attributes (defined in a Report Structure) can be attached.

The first step in defining a metadata structure definition is to define the target identifiers. Target identifiers are used by metadata report structures (which will be described later) to define how a report instance in a metadata set identifies the object it is providing reference metadata for. There are two type of target identifiers; full and partial. Both of these target identifiers consist of individual identifier components.

3.2.2 Identifier Components

A target identifier in a metadata structure definition consists of one or more identifier components. These individual identifier components are combined in the target identifier to describe the object or cross section of objects which reference metadata describes. This is much the same principle as on the data side where the individual Dimensions serve to identify the observation value, except that for metadata the artefact that is identified can be any artefact in the SDMX information model (it is not restricted, as on the data side, to the observation value).

Creating an identifier component begins with assigning it an identification that is unique within the metadata structure definition.

An identifier component must have an association to an identifiable object type - which must be one or the formally recognized object classes from the information model. For example, an identifier component may have an association to a category. In this case, the identifier component will reference a category in a reference metadata report. By defining this object association, a metadata structure definition is able to describe which types of objects the metadata reports can describe.

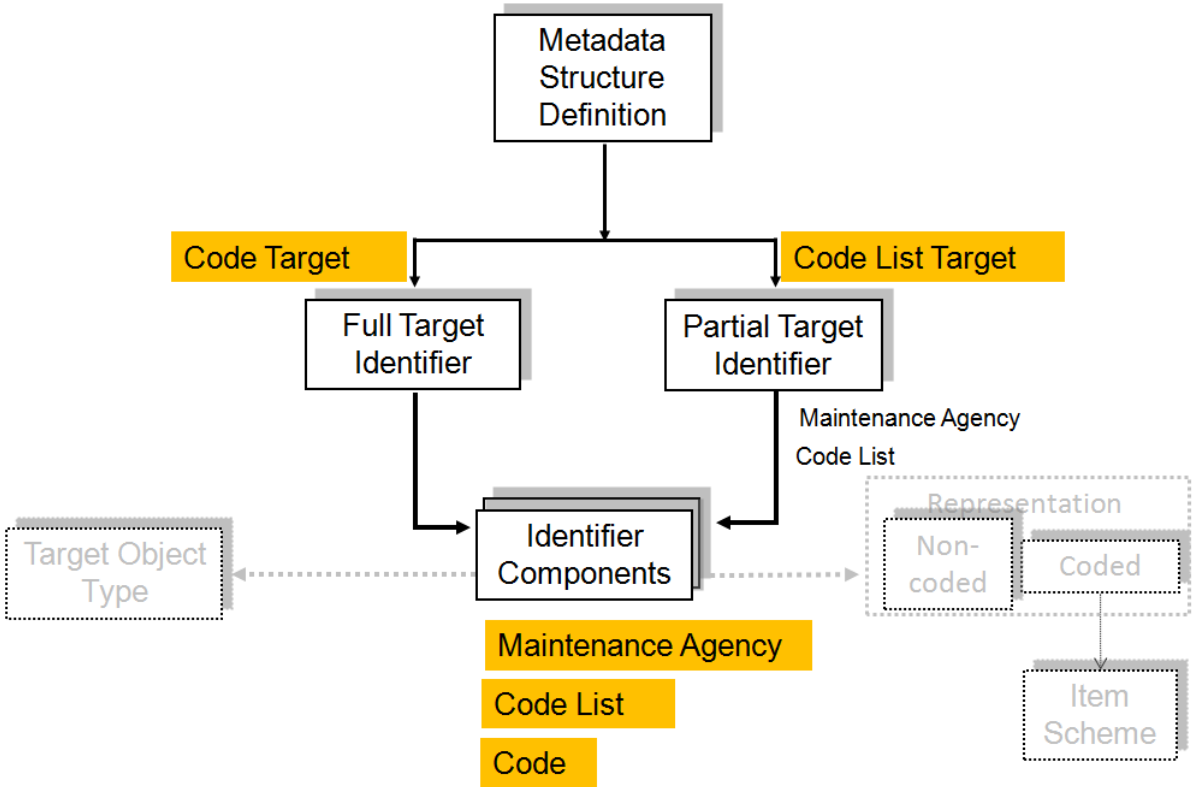
Finally, an identifier component can have an association to an item scheme which enumerates the possible values for the component. This effectively allows a metadata structure definition to restrict the values (identifiers) of the objects which the reference metadata may describe. An item scheme is a generic name in the information model whose specific types are: code list, concept scheme, organisation scheme, category scheme. For example, if an identifier component has an associated object type of a data provider it may be given a representation scheme of a particular organisation scheme. In this scenario, reference metadata conforming to the metadata structure definition would only be allowed to attach to data providers contained in the organisation scheme.

3.2.3 Full and Partial Target Identifiers

The Full and Partial Target Identifiers serve much the same purpose as the Key and Group Key for data. In a Data Structure Definition, the Key specifies the Dimensions and their allowable format and, for a coded format, the identity if the code list. The Group Key identifies a sub-set of the Dimensions to which Data Attributes can be attached. Similarly, the Full Target Identifier specifies all of the Identifier Components in the scope of this MSD, whilst the Partial Target Identifier identifies a sub-set of the Identifier Components to which Metadata Attributes may be attached: it will be seen later that the metadata Report Structure actually specifies the Metadata Attributes and it is this Report Structure that identifies the Full or Partial Target Identifier to which its Metadata Attributes are attached.

It is important to understand that each Partial Target Identifier must contain only those Identifier Components required to identify a **single** SDMX artefact or cross section of artefacts. For example, if it is required to attach metadata to either a Code or a Code List, then there must be two Target Identifiers specified, one that contains the Identifier Components for a Code List (the Maintenance Agency and the Code List identifier) and one that contains the Identifier Components for a Code (the Maintenance Agency, the Code List identifier, and the Code identifier).

As both the Full Target Identifier and the Partial Target Identifier can be used to identify a single SDMX artefact, the example above (or Code List and Code) could be achieved by defining a Full Target Identifier with three Identifier Components (Maintenance Agency, Code List identifier, Code identifier) with the Partial Target Identifier containing just the Identifier Components of Maintenance Agency and Code List identifier. This is shown in the schematic below.



In this example, the Full Target Identifier is used by the Report Structure for which its Metadata Attributes are to be attached to a Code and the Partial Target Identifier is used by the Report Structure for which its Metadata Attributes are to be attached to a Code List.

It is permissible to specify a Full Target Identifier that is not used by a Report Structure. Indeed, this is quite common: in these cases the only purpose of the Full Target Identifier is to define the Identifier Components (and importantly the object type and item scheme used) so that they may be included in Partial Target Identifiers.

3.2.4 Demonstration

At this point, a sample metadata structure definition will be built in the SDMX Metadata Editor in order to enhance the user's understanding. The example will be based on the ESMS cited earlier in this document. The first step will be to create the metadata structure definition and to define the Full Target Identifier. In Figure Figure 2 of the ESMS sample cited above, it can be seen that the reference metadata is attached at the level of a classification. In addition, it should also be possible to also attach the reference metadata to a specific data provider. Based on this, we can define a Full Target Identifier.

The annotated schematic below shows the content of the Full Target Identifier.

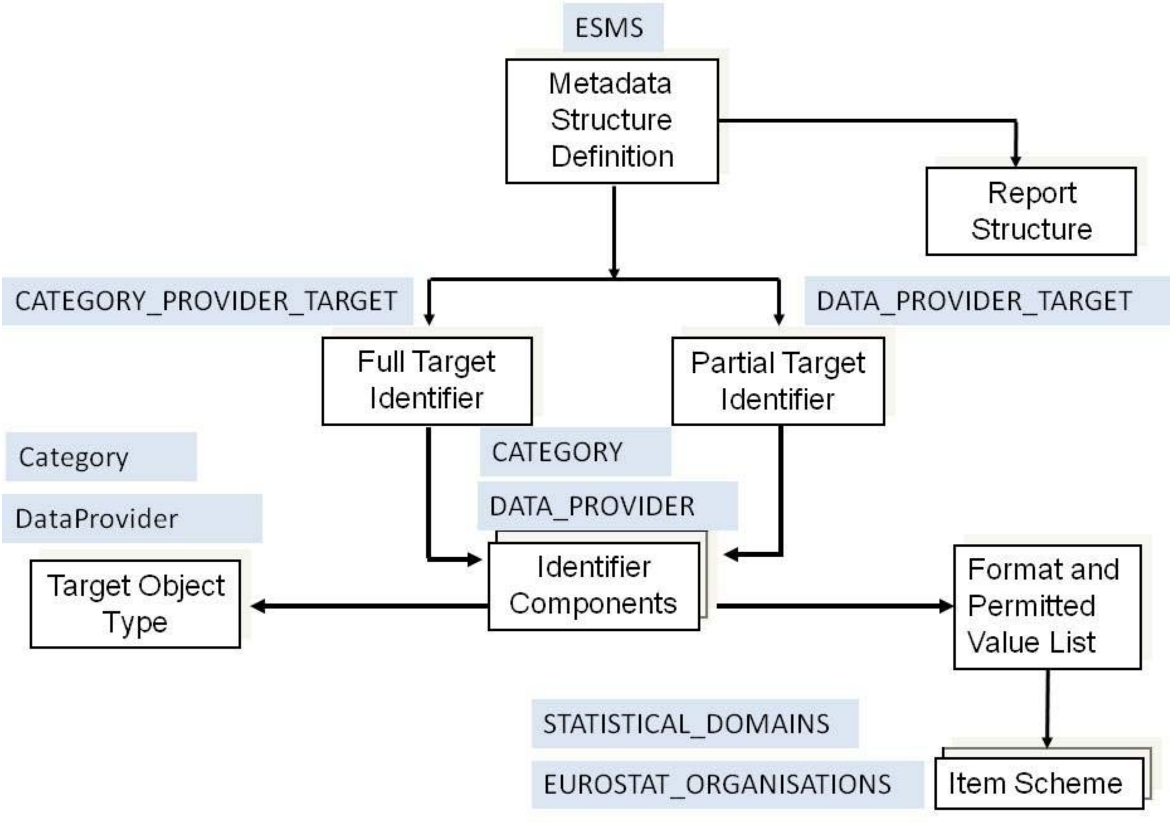


Figure 6: Schematic of Content of Target Identifiers

An example of the specification of this using the MSD Editor is shown in FigureFigure 7 below.

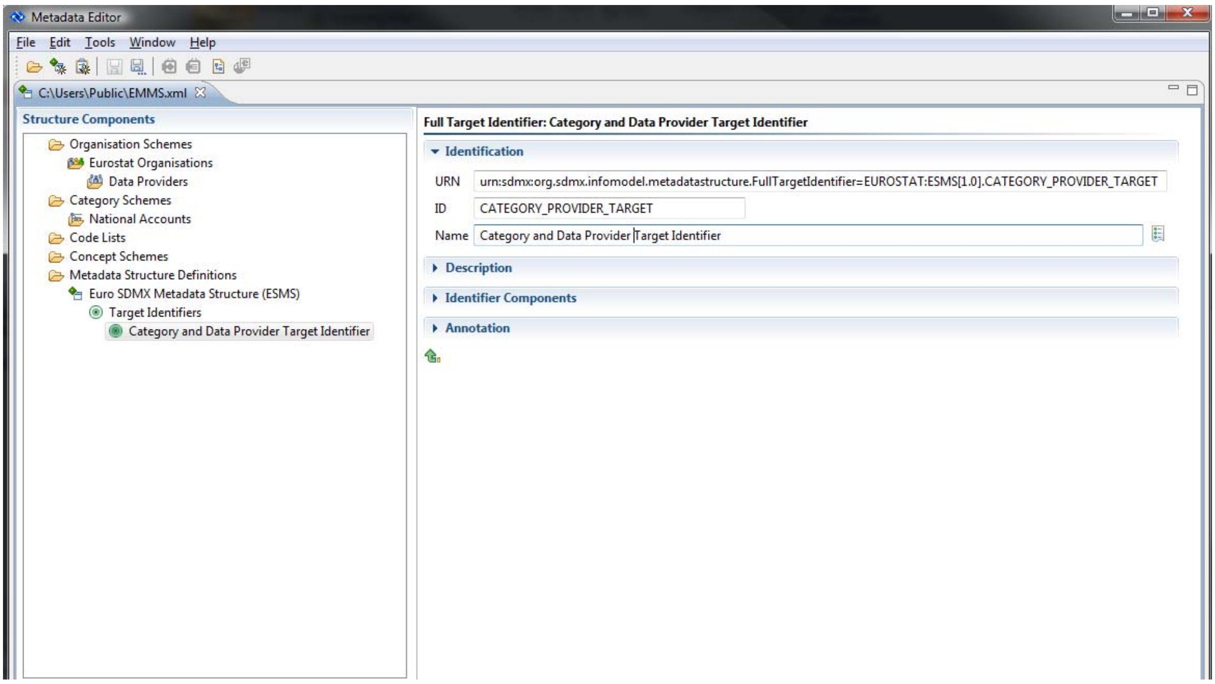


Figure 7: Creation of full target identifier in the SDMX Metadata Editor

Once the Full Target Identifier is defined, its identifier components must be added. Two Identifier Components have been identified for the ESMS, one identifying a Category and another identifying a Data Provider. Identifier Components are typically assigned descriptive identifiers. For the ESMS example, the two Identifier Components are identified as 'CATEGORY' and 'DATA_PROVIDER'. As described above, these components must have a target object associated with them. These classes are 'Category' and 'DataProvider', respectively.

Finally, the Identifier Components may be assigned an Item Scheme which defines their representation. For the CATEGORY Identifier Component, this representation will be a Category Scheme. For the DATA_PROVIDER Identifier Component, an Organisation Scheme is used for the representation. Note that in the left navigation panel in Figure Figure 7 that a Category Scheme and Organisation Scheme have been defined. For the purpose of this exercise, these do not need to contain any details. These schemes will be used for the representation of the Identifier Components. Figure Figure 8 shows an example of the completed category Identifier Component.

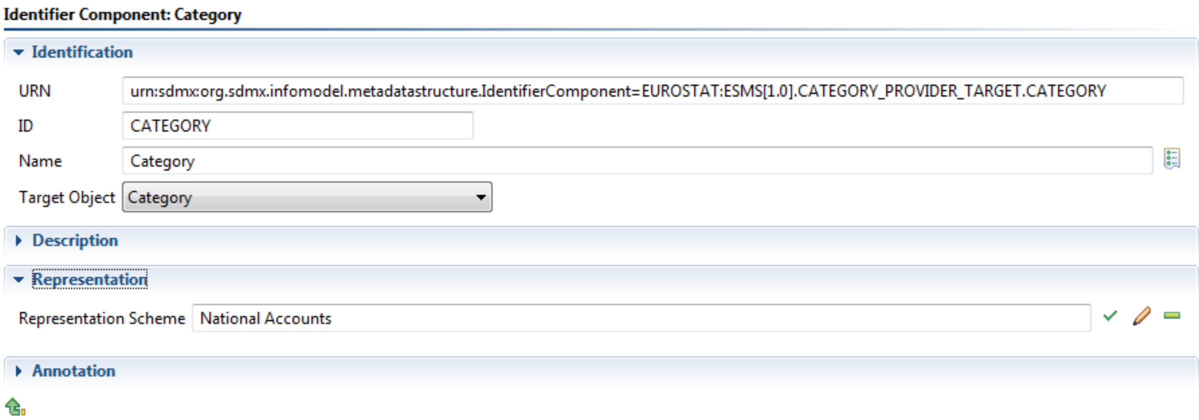


Figure 8: Creation of ESMS CATEGORY identifier component in the SDMX Metadata Editor

Now that the Full Target Identifier has been identified, reference metadata can be attached to data reported by specific data provider against a specific category. The data provider can be completed in a similar fashion.

For the purpose of this demonstration, even though it is not shown in the ESMS example, a Partial Target Identifier can also be defined. Suppose, for example, that some reference metadata does not vary from category to category - that is to say that is the same for a given data provider. In a data structure definition we would define a group to attach such a data attribute, and so in a metadata structure definition we will define a Partial Target Identifier. A partial target identifier, 'DATA_PROVIDER_TARGET', can be defined. This Partial Target Identifier only utilizes the DATA_PROVIDER Identifier Component from the Full Target Identifier. Notice that when defining a Partial Target Identifier, the Identifier Component is not defined; rather it is referenced – using one of the Identifier Components in the Full Target Identifier. The figure below shows the partial target definition.

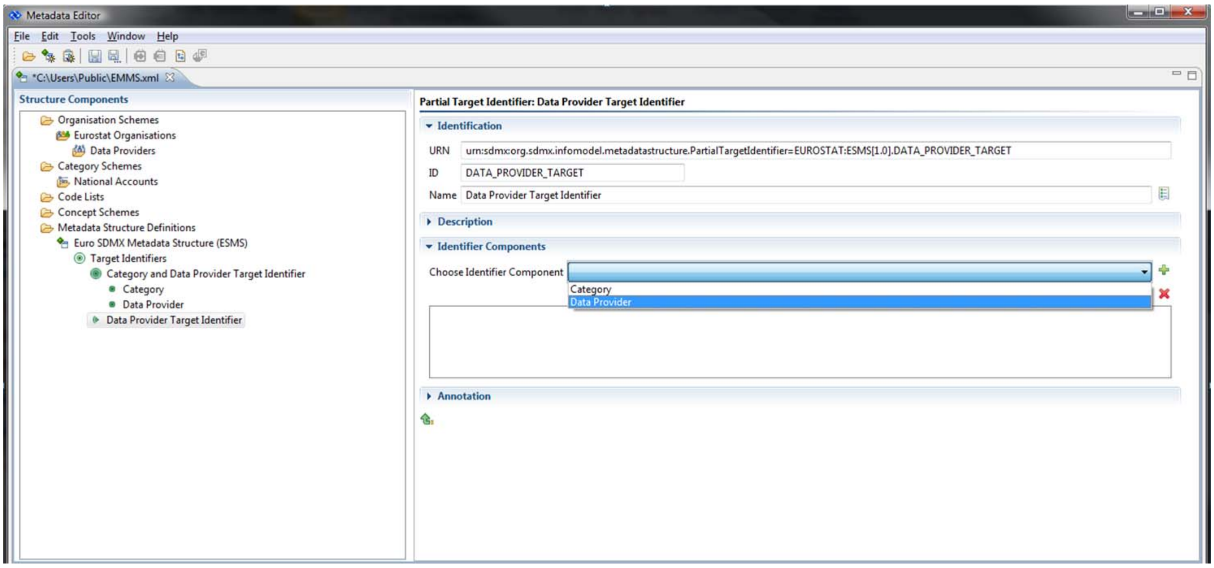


Figure 9: ESMS partial target identifier definition

3.2.5 Sample

The Figure below shows an example of the SDMX-ML output of target identifiers that were defined.


```
- <MetadataStructureDefinitions>
  - <structure:MetadataStructureDefinition id="ESMS" agencyID="EUROSTAT" version="1.0"
      urn="urn:sdmx:org.sdmx.infomodel.metadatastructure.MetadataStructureDefinition=EUROSTAT:ESMS[1.0]">
      <structure:Name>Euro SDMX Metadata Structure (ESMS)</structure:Name>
    - <structure:TargetIdentifiers>
      - <structure:FullTargetIdentifier id="PROVIDER_AND_FLOW_TARGET"
          urn="urn:sdmx:org.sdmx.infomodel.metadatastructure.FullTargetIdentifier=EUROSTAT:ESMS
          [1.0].PROVIDER_AND_FLOW_TARGET">
          <structure:Name>Data Provider and Dataflow Target Identifier</structure:Name>
        - <structure:IdentifierComponent id="DATAFLOW"
            urn="urn:sdmx:org.sdmx.infomodel.metadatastructure.IdentifierComponent=EUROSTAT:ESMS
            [1.0].PROVIDER_AND_FLOW_TARGET.DATAFLOW">
            <structure:Name>Dataflow</structure:Name>
            <structure:TargetObjectClass>DataFlow</structure:TargetObjectClass>
            <structure:RepresentationScheme representationSchemeType="Category" representationScheme="NATIONAL_ACCOUNTS"
              representationSchemeAgency="EUROSTAT" />
          </structure:IdentifierComponent>
        - <structure:IdentifierComponent id="DATA_PROVIDER"
            urn="urn:sdmx:org.sdmx.infomodel.metadatastructure.IdentifierComponent=EUROSTAT:ESMS
            [1.0].PROVIDER_AND_FLOW_TARGET.DATA_PROVIDER">
            <structure:Name>Data Provider</structure:Name>
            <structure:TargetObjectClass>DataProvider</structure:TargetObjectClass>
            <structure:RepresentationScheme representationSchemeType="Organisation"
              representationScheme="EUROSTAT_ORGANISATIONS" representationSchemeAgency="EUROSTAT" />
          </structure:IdentifierComponent>
        </structure:FullTargetIdentifier>
      - <structure:PartialTargetIdentifier id="DATA_PROVIDER_TARGET"
          urn="urn:sdmx:org.sdmx.infomodel.metadatastructure.PartialTargetIdentifier=EUROSTAT:ESMS
          [1.0].DATA_PROVIDER_TARGET">
          <structure:Name>Data Provider Target</structure:Name>
          <structure:IdentifierComponentRef>DATA_PROVIDER</structure:IdentifierComponentRef>
        </structure:PartialTargetIdentifier>
      </structure:TargetIdentifiers>
```


Figure 10: ESMS target identifiers SDMX-ML sample

Note that in the Full Target Identifier, the Identifier Component definition contains the target object class and the representation scheme. However, in the Partial Target Identifier the DATA_PROVIDER Identifier Component is only referenced by its identification. Implementation Limitations

The metadata structure definition is a very flexible mechanism that supports the specification of any type of report that may be required for the objects in the information model. However, there are some limitations that the user needs to be aware of.

1. Whilst it is theoretically possible to define an MSD that can identify specific regions of a data set in terms of the series keys of that region, in practice this is very difficult to specify in a generic manner. Therefore, in SDMX-ML version 2.0 the user is really limited to defining Metadata Structure Definitions which allow for reference metadata to be attached to data conforming to a particular Key Family definition. This is achieved by defining Identifier Components which have identifiers matching the dimension concepts, thus allowing the Full and Partial Target Identifiers to contain full or partial data keys.

2. There are restrictions on the permitted character set that can be used to identify objects in a generic metadata report. This is limited to the character set used by the object identifiers (in XML technical terms this is limited to XML schema NMTOKEN type). Therefore, whilst in SDMX it is possible to identify any object by its URN, it is invalid to place a URN in the value reported for an Identifier Component.

3. For users wishing to use metadata structure specific schemas, be aware that

   .the metadata structure definition specific schema binding rules (i.e. the rules that define how metadata structure definition specific schemas are to be created) do not fully describe how item schemes such as category schemes or organisation schemes should be used to create enumerations. Further, the schema binding rules do not address how identifier components without a specific representation should be created.

3.3 Report Structure

3.3.1 Schematic

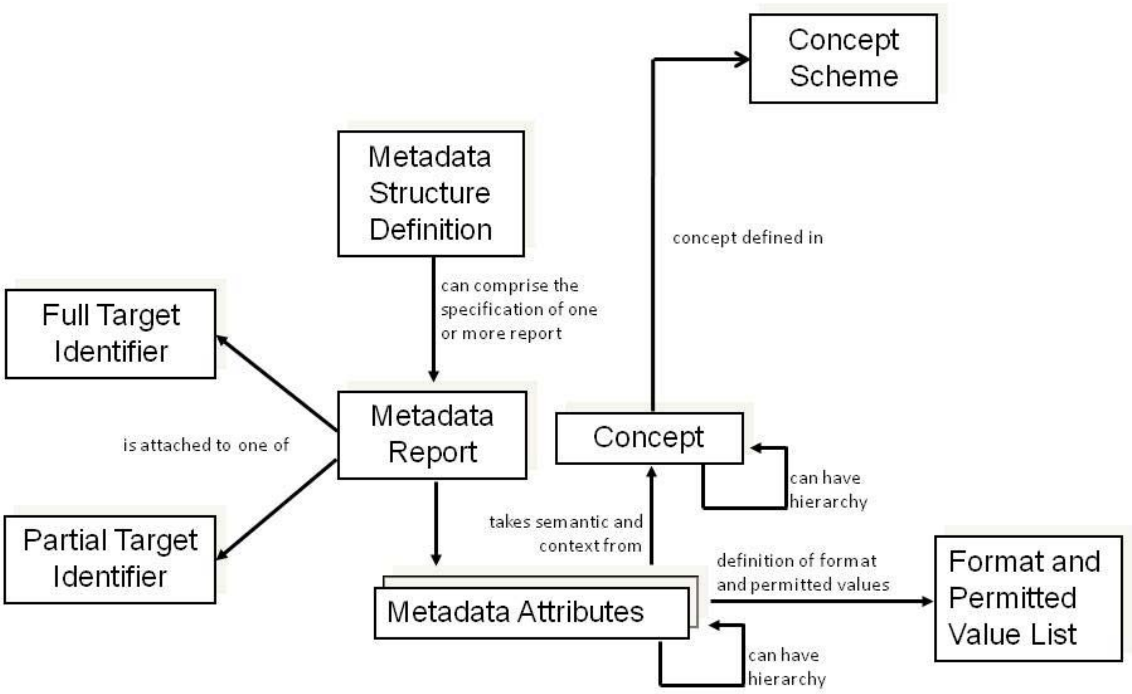


Figure 11: Schematic of the Information Model for a Report Structure

The next step in defining a metadata structure definition is defining the structure of the reference metadata reports. This structure in described in one or more report structures. A Report Structure consists of Metadata Attributes arranged in a hierarchy. In addition, each Report Structure is associated with a target identifier. The following sections will describe a Report Structure definition, its association to a target identifier, and the Metadata Attributes which comprise the report.

3.3.2 Report Structure Definition

A Report Structure must be given identification. This identification must be unique within a Metadata Structure Definition. The content of a Report Structure are Metadata Attributes. A Metadata Attribute can comprise child Metadata Attributes. In this way it is possible to define hierarchical structures of Metadata Attributes. Hierarchies of this nature are quite common in data quality frameworks such as the ESMS. A hierarchy can also be useful to group Metadata Attributes that comprise a single concept such as a Contact, whose child Metadata Attributes would be the individual concepts used to define a contact such as name, address, organisation, e-mail address etc.

The details of these metadata attribute definitions will be described in more detail below.

3.3.3 Target

Each Report Structure must have an associated target identifier (either a Full Target Identifier or a Partial Target Identifier) which defines the objects which the report can be attached in the corresponding metadata sets. Since each Report Structure can only have one target identifier, it is necessary to duplicate Report Structures if the Metadata Attributes in a report can be attached to the object or intersection of objects described by more than one target identifier.

3.3.4 Metadata Attributes

A Metadata Attribute takes its semantic from a Concept, and therefore has a mandatory association to a Concept. The Metadata Attribute can have a value when reported in a Metadata Set and/or child Metadata Attributes. These child Metadata Attributes allow the reference metadata report to have a hierarchical structure. Within the context of this structure, Metadata Attributes can be declared as mandatory or optional. Finally, when a Metadata Attribute is specified as containing a value (when reported in a Metadata Set), a representation can be defined. As with a Dimension or a Data Attribute in a DSD, the Metadata Attribute can take a coded representation from a code list or define an un-coded text format.

3.3.5 Demonstration

Expanding on the ESMS example, a Report Structure will be defined in SDMX Metadata Editor. The first step is to define a Report Structure, giving it identification and associating it with a target identifier. Since the ESMS example that is being referenced is attached to both a Category and a Data Provider, the Report Structure should reference the Full Target Identifier that was created in the previous example.

The annotated schematic below shows the content of the Report Structure.

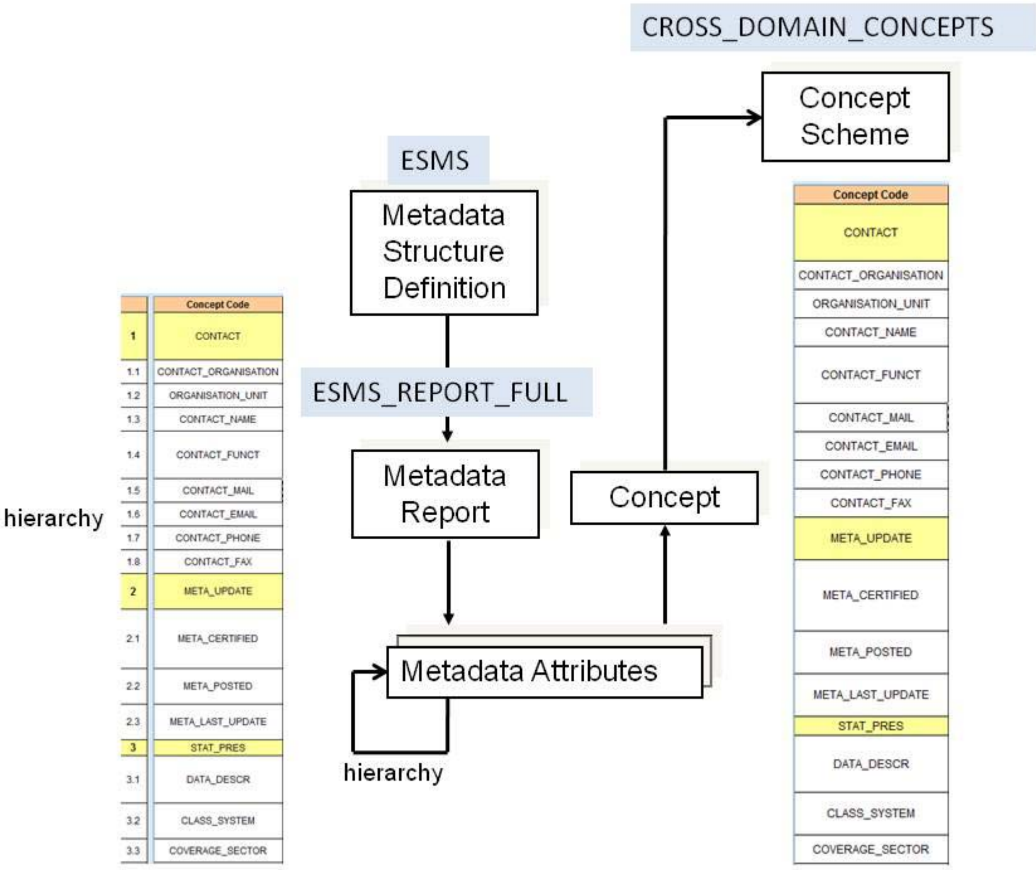


Figure 12: Schematic of content of the Report Structure

Note that the Concept Scheme contains a flat list of Concepts. These can be built into a hierarchy when used in Metadata Attributes in the Report Structure. Therefore, in this example the hierarchy defined in the ESMS can be faithfully represented in the MSD.

The figure below shows an example of the Report Structure definition.

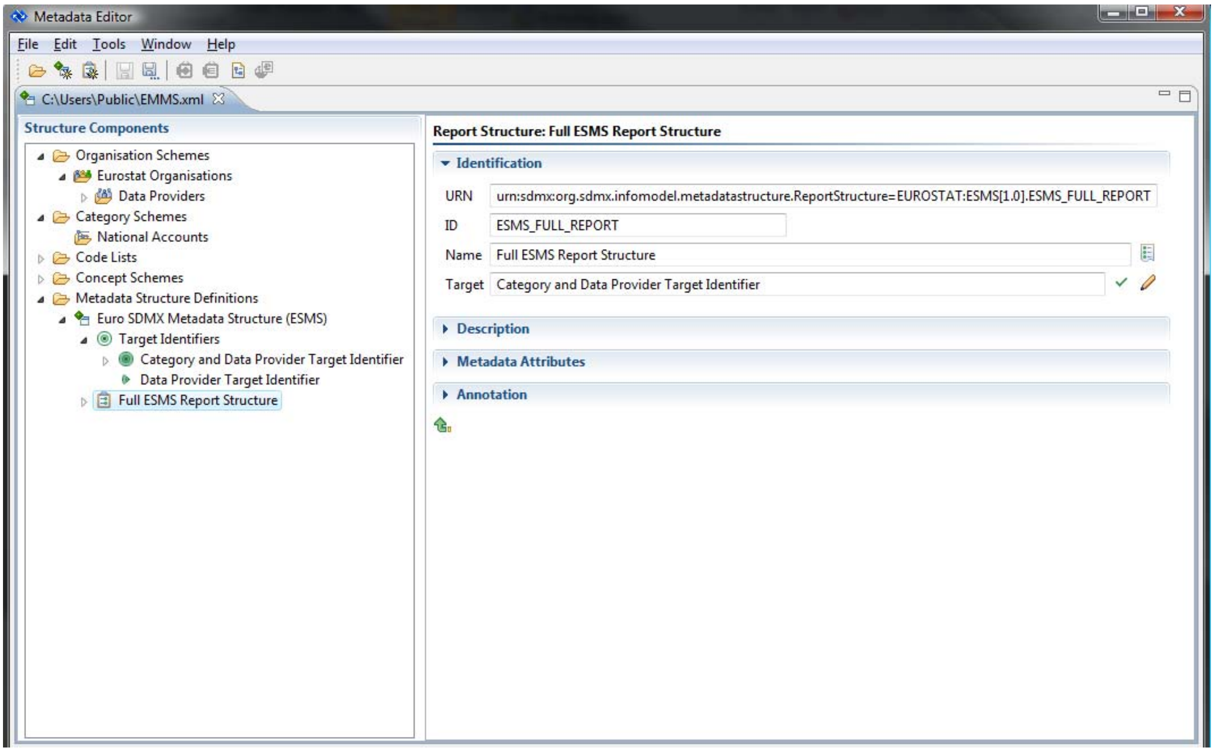


Figure 13: - ESMS full report structure definition example.

Examining the sample report in Figure Figure 4, it can be seen that the metadata report contains Concepts such as a contact and information about when the metadata was updated. These Metadata Attributes are added as direct children of the Report Structure. For now, only the concept reference will be defined. Note that in the left navigation panel in the figure above there is a Concept Scheme, ‘Eurostat Cross Domain Concepts’, defined. The Concepts defined in this scheme are used in the definition of the Metadata Attributes.

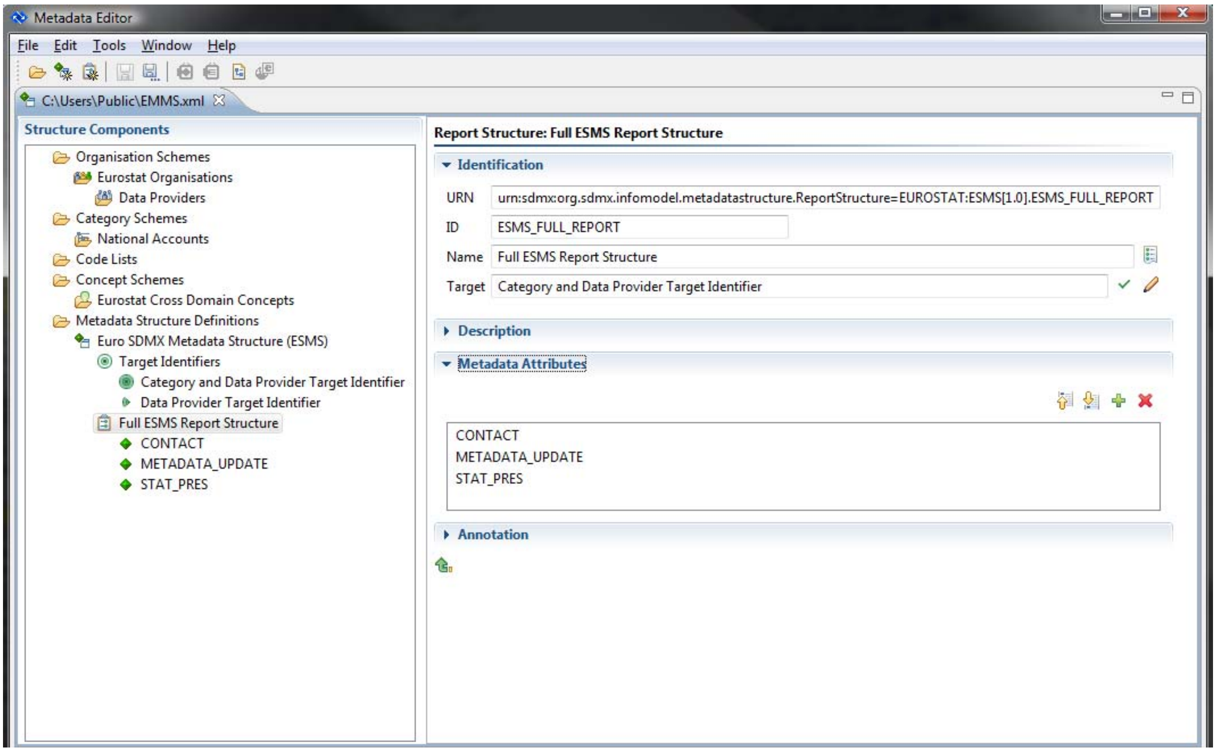


Figure 14: ESMS contact, metadata update, and statistical presentation attribute definitions.

After the Metadata Attributes are added to the Report Structure, their details must be defined. Considering the ESMS example, one may require that contact information always be provided. Therefore it should be defined as being mandatory. As for the content of the contact, it can be seen that it has sub-structure, including the contact organisation and organisation unit. Based on what is known about typical contact details, other child metadata attributes can also be assumed. Figure 10 shows the details of a full contact metadata attribute hierarchy definition.

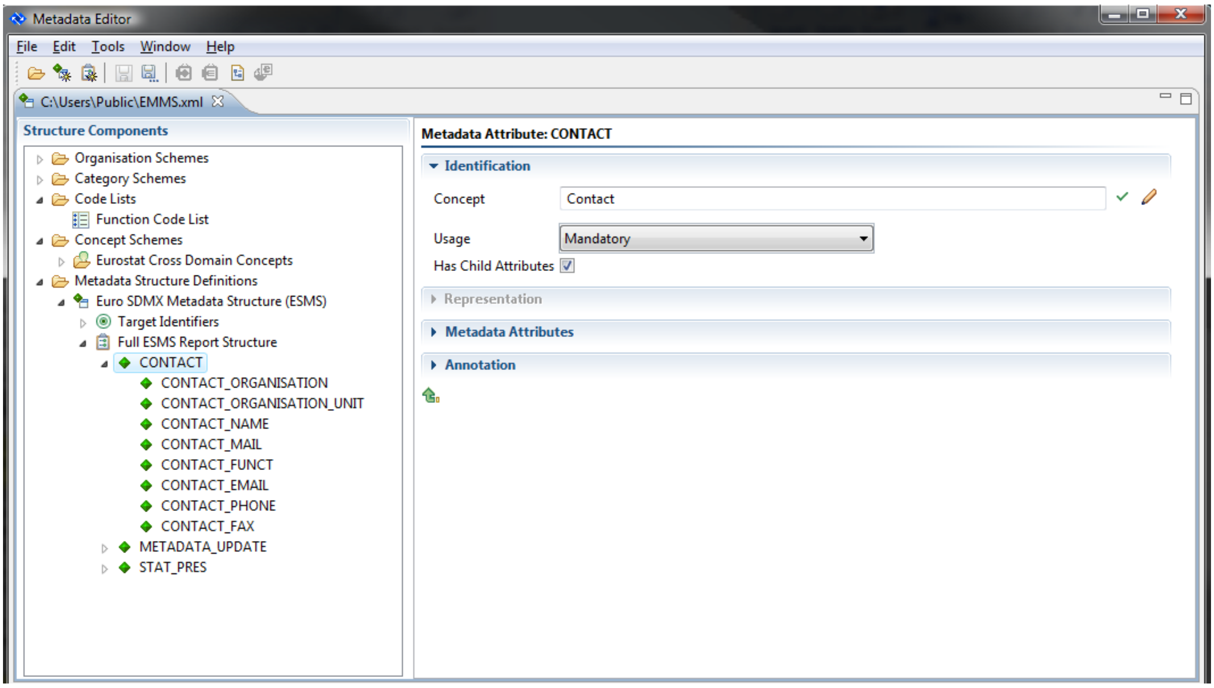


Figure 15: ESMS contact metadata attribute definition details.

Note that in the left panel you can see that the hierarchy of the CONTACT Metadata Attribute. If the details of the CONTACT_NAME Metadata Attribute are examined, it can be seen that it supplies a text format specification for its value (as well as the fact that its usage is conditional).

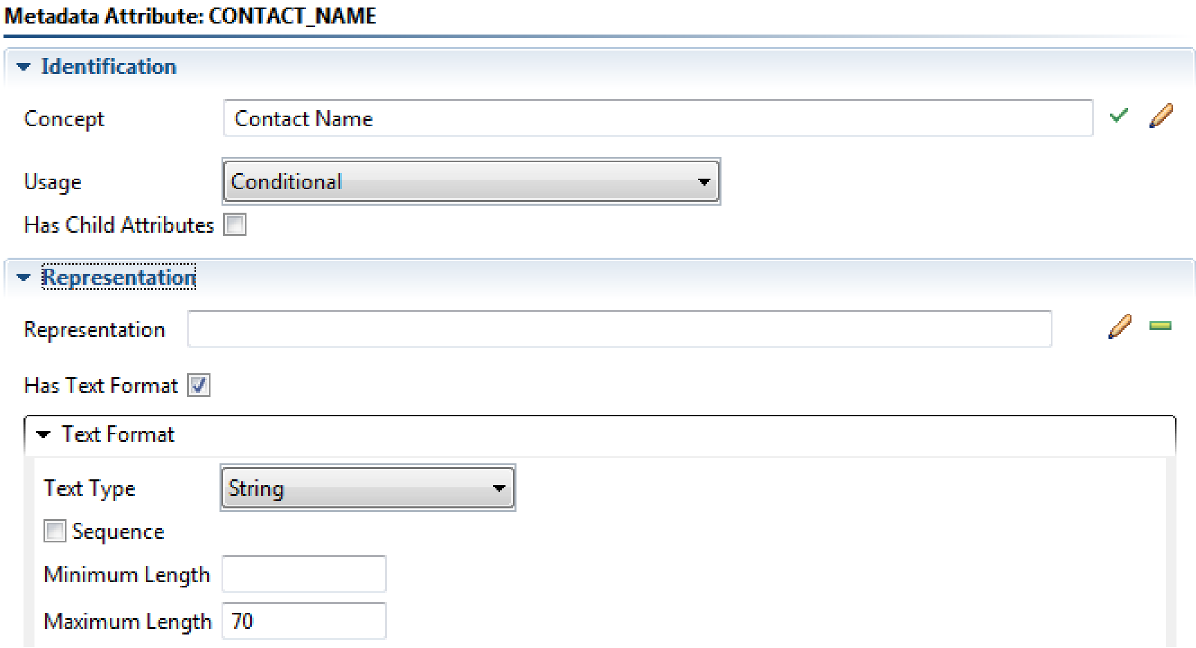


Figure 16: ESMS CONTACT_NAME metadata attribute definition details.

Contrast this to the CONTACT_FUNCT metadata attribute, where the representation is taken from a Code List.

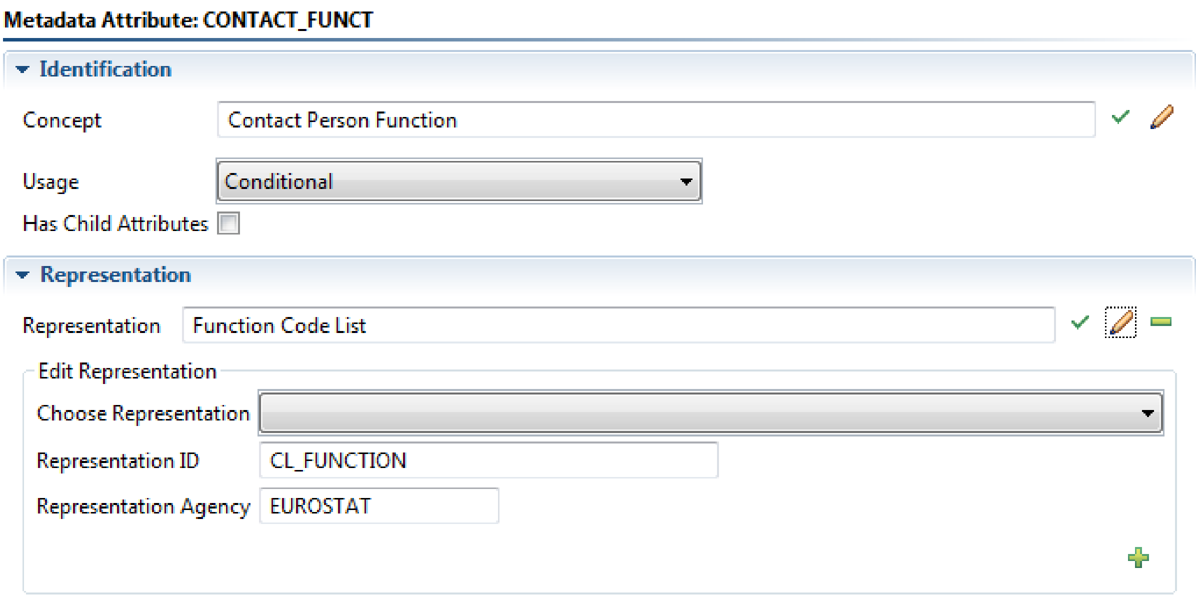


Figure 17: ESMS CONTACT_FUNCT metadata attribute definition details.

A Metadata Structure Definition may contain multiple Report Structures. For the sake of demonstration, suppose that the contact information for some Data Providers does not vary by Category. In this scenario a second Report Structure can be defined, and associated with the Partial Target Identifier for a Data Provider defined in the previous section. The details of the

CONTACT Metadata Attribute would not be changed. By defining this second Report Structure, one now has the option of providing contact information for all data sets from a Data Provider or of providing it as part of a complete report against data sets for a specific Category and Data Provider.

1.1.1 Sample

FigureFigure 18 shows a sample of the full Report Structure that was defined in the previous section.


```
- <structure:ReportStructure id="ESMS_FULL_REPORT" target="CATEGORY_PROVIDER_TARGET"
   urn="urn:sdmx:org.sdmx.infomodel.metadatastructure.ReportStructure=EUROSTAT:ESMS[1.0].ESMS_FULL_REPORT">
   <structure:Name xml:lang="en">Full ESMS Report Structure</structure:Name>
 - <structure:MetadataAttribute conceptRef="CONTACT" usageStatus="Mandatory" conceptSchemeRef="CROSS_DOMAIN_CONCEPTS"
     conceptSchemeAgency="EUROSTAT" conceptVersion="1.0">
   - <structure:MetadataAttribute conceptRef="CONTACT_ORGANISATION" usageStatus="Conditional"
       conceptSchemeRef="CROSS_DOMAIN_CONCEPTS" conceptSchemeAgency="EUROSTAT" conceptVersion="1.0">
       <structure:TextFormat textType="String" maxLength="70" />
     </structure:MetadataAttribute>
   - <structure:MetadataAttribute conceptRef="CONTACT_ORGANISATION_UNIT" usageStatus="Conditional"
       conceptSchemeRef="CROSS_DOMAIN_CONCEPTS" conceptSchemeAgency="EUROSTAT" conceptVersion="1.0">
       <structure:TextFormat textType="String" maxLength="70" />
     </structure:MetadataAttribute>
   - <structure:MetadataAttribute conceptRef="CONTACT_NAME" usageStatus="Conditional"
       conceptSchemeRef="CROSS_DOMAIN_CONCEPTS" conceptSchemeAgency="EUROSTAT" conceptVersion="1.0">
       <structure:TextFormat textType="String" maxLength="70" />
     </structure:MetadataAttribute>
   - <structure:MetadataAttribute conceptRef="CONTACT_MAIL" usageStatus="Conditional"
       conceptSchemeRef="CROSS_DOMAIN_CONCEPTS" conceptSchemeAgency="EUROSTAT" conceptVersion="1.0">
       <structure:TextFormat textType="String" />
     </structure:MetadataAttribute>
     <structure:MetadataAttribute conceptRef="CONTACT_FUNCT" usageStatus="Conditional"
       conceptSchemeRef="CROSS_DOMAIN_CONCEPTS" conceptSchemeAgency="EUROSTAT" conceptVersion="1.0"
       representationScheme="CL_FUNCTION" representationSchemeAgency="EUROSTAT" />
   - <structure:MetadataAttribute conceptRef="CONTACT_EMAIL" usageStatus="Conditional"
       conceptSchemeRef="CROSS_DOMAIN_CONCEPTS" conceptSchemeAgency="EUROSTAT" conceptVersion="1.0">
       <structure:TextFormat textType="String" />
     </structure:MetadataAttribute>
   - <structure:MetadataAttribute conceptRef="CONTACT_PHONE" usageStatus="Conditional"
       conceptSchemeRef="CROSS_DOMAIN_CONCEPTS" conceptSchemeAgency="EUROSTAT" conceptVersion="1.0">
       <structure:TextFormat textType="String" />
     </structure:MetadataAttribute>
   - <structure:MetadataAttribute conceptRef="CONTACT_FAX" usageStatus="Conditional"
       conceptSchemeRef="CROSS_DOMAIN_CONCEPTS" conceptSchemeAgency="EUROSTAT" conceptVersion="1.0">
       <structure:TextFormat textType="String" />
     </structure:MetadataAttribute>
   </structure:MetadataAttribute>
 + <structure:MetadataAttribute conceptRef="METADATA_UPDATE" usageStatus="Conditional"
     conceptSchemeRef="CROSS_DOMAIN_CONCEPTS" conceptSchemeAgency="EUROSTAT" conceptVersion="1.0">
 + <structure:MetadataAttribute conceptRef="STAT_PRES" usageStatus="Conditional" conceptSchemeRef="CROSS_DOMAIN_CONCEPTS"
     conceptSchemeAgency="EUROSTAT" conceptVersion="1.0">
 </structure:ReportStructure>
```


Figure 18: ESMS report structure SDMX-ML sample.

Note that the hierarchy is apparent by the nesting of the Metadata Attributes.

3.3.6 Implementation Limitations

As with the target identifiers, there are some limitations as to what the SDMX-ML 2.0 implementation of the report structure can support.

1. Although the generic metadata report allows for Metadata Attributes to contain both a value and child Metadata Attributes, the metadata structure definition specific schema bindings do not allow this. For example, if you wish to allow a value to be entered for the CONTACT in the above example as well as allowing child Metadata Attributes you will need to use the generic metadata set. Note that the tools used in this demonstration only allowed for child representation or a value, but not both.

2. The Metadata Structure Definition specific schema binding has a restriction on the definition of the Metadata Attributes. The schema binding rules dictate that a global complex type must be created for each metadata attribute defined in the metadata

structure definition. This result of this restriction is that if a Concept is used more than once it must always have the same content. For example, if a Metadata Structure Definition contained two report structures each containing a CONTACT metadata attribute - it would not be possible for one the to include a child ADDRESS Metadata Attribute whereas the other one excluded it.

4 Metadata Sets

4.1 Scope of this Chapter

Whilst it is not the purpose of this book to describe Metadata Sets, nevertheless it is important to understand the relationship between the Metadata Set and the Metadata Structure Definition and in particular the derivation of Metadata Structure Definition specific schemas. This chapter describes briefly these issues.

4.2 Generic Metadata Set

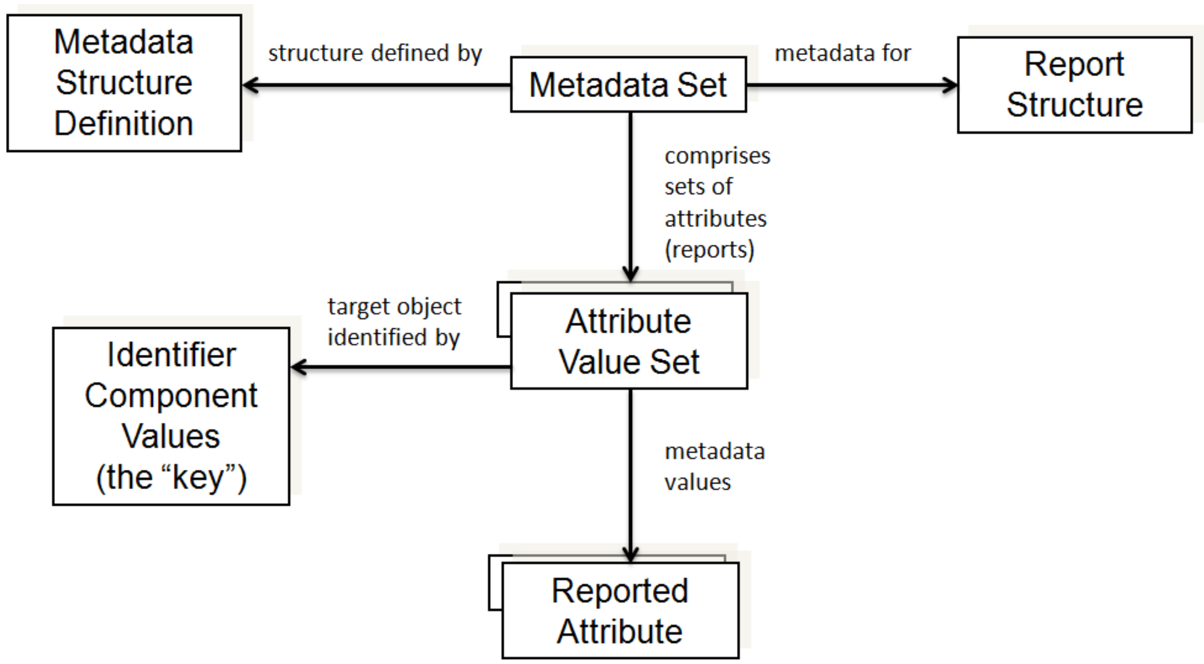


Figure 19: Schematic of the Generic Metadata Set

Just as data has a structural metadata neutral format for exchanging data, reference metadata has the generic metadata messages. Reference metadata in a generic metadata set must conform to single report structure in a metadata structure definition. The report is organised in attribute value sets, where each set identifies the target object it is describing as well as the details of the metadata attributes contained in the report. A generic metadata set allows for multiple attribute value sets, therefore it is possible to have multiple report instances in a single metadata set, providing that all reports conform to the same report structure. As with the generic data structure, this format provides no built in validation against the metadata structure definition, and is therefore useful when exchanging reference metadata between systems that are not capable of processing metadata structure definition specific reference metadata.

4.3 Metadata Structure Definition Specific Metadata Set

SDMX also provides for a reference metadata format that is similar in concept to the Utility data format. This format is referred to in the SDMX-ML schemas as the metadata report format. Just like the Utility data format, the base SDMX-ML schemas only define a stub for the reference metadata report, which is refined via extensions and substitutions in a XML schema created from the metadata structure definition.

4.3.1 Schema

Part 03A of the SDMX specification (SDMX-ML: Schema and Documentation) defines the rules for creating a metadata structure definition specific XML schema in section 6.6. The purpose of this section is not to restate the rules defined in this section, but rather to give an overview of what the net effect of these rules are. Like the Utility data format, the schema derived from the metadata structure definition provides strict validation of reference metadata against the structure described in the metadata structure definitions. For example, if a given report structure is used the complete target identification must be provided. Further, if the metadata structure definition defines an enumerated representation for an identifier component, the metadata structure definition specific XML schema will enforce that only the values from the reference items scheme can be used. Similarly, if a metadata attribute is required in a report structure, then the metadata structure definition specific schema will require that it is present in the reference metadata report.

4.3.2 Instance

A metadata structure definition specific metadata report instance differs quite substantially from the generic instance. In the generic instance the components are referred to by elements bearing the general object name such as ‘component value’ or ‘reported attribute’. In the specific metadata report instance the XML elements are given names based on the identification of the component. In the ESMS example, the CONTACT metadata attribute would be contained in an XML element named CONTACT. This is similar to the Utility format where dimensions are contained in elements with the name of the dimension's concept identifier.

4.4 Interacting Between Formats

In theory, it should be possible to derive reference metadata instances in one of the above formats from the other format. However, the biggest difference when interacting between the generic and the metadata structure definition specific formats is that the generic format only allows for reference metadata to be provided against a single report structure, whereas the metadata structure definition specific format allows for multiple reports against any report structure format. This means that it is not always possible for a single metadata structure definition specific reference metadata instance to be transformed to a single generic reference metadata instance. Another difference (highlighted earlier) is that the generic format allows for a metadata attribute to contain both a value and child attributes, where as the metadata structure definition specific format only allows for one or the other. Because of these limitations, one must take care in defining a metadata structure definition, and attempt to understand how it will be used.

5 Glossary

Table 2 presents the list of concepts and acronyms with their definition.

Concept	Definition
DSD	Data Structure Definition
ESMS	Euro SDMX Metadata Structure
MSD	Metadata Structure Definition
SDMX	Statistical Data and Metadata eXchange.
SDMX-IM	SDMX Information Model
SDMX-ML	SDMX Markup Language - XML format for the exchange of SDMX-structured data and metadata
XML	eXtensible Markup Language

Table 2 - Glossary