Metadata Working Group

Luxembourg, 7-8 June 2007

BECH Building, Room Ampère

ITEM 5 OF THE AGENDA

Reference Metadata Support in SDMX Version 2.0

Reference Metadata Support in SDMX Version 2.0

Author: Chris Nelson, Standards Expert
08 June 2007

Contents

Scope

The scope of this paper is to outline the support for metadata in the SDMX version 2.0 standards. Metadata, in its various forms, is at the heart of the SDMX standards, and much of the standard is concerned with metadata. This paper concentrates on Reference Metadata[1] but, as this type of metadata usually describes data or the provision/publishing of data, the paper also discusses these aspects.

From Version 1.0 to Version 2.0

Version 1.0 of the SDMX standards (ISO/TS 17369) is concerned mainly with data sets and the structural definitions for data sets. The diagram below shows the main artefacts supported by the version 1.0 standards.
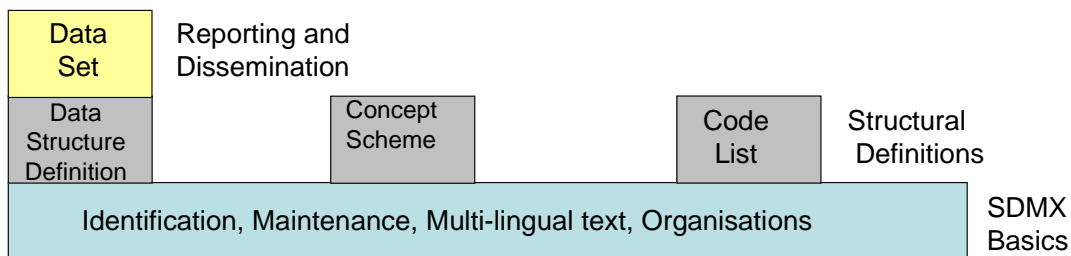
Version 2.0 has seen a large increase in the functionality supported by the SDMX standards. The additional functionality has resulted largely from requirements requested, both during the development of the version 1.0 standard and from early adopters of the version 1.0 standard. The version 2.0 standards introduced support for reference metadata (metadata structure definition, metadata set), mapping between structural metadata (e.g. code list maps, data and metadata structure maps, category and concept scheme maps), Category Scheme (e.g. this supports a domain category scheme), hierarchic code lists, transformation metadata (to better support primary reporting), and registry based functionality (data and metadata provisioning, subscription and notification, data and metadata registration, and data and metadata discovery). The diagram below shows the main artefacts supported at version 2.0.
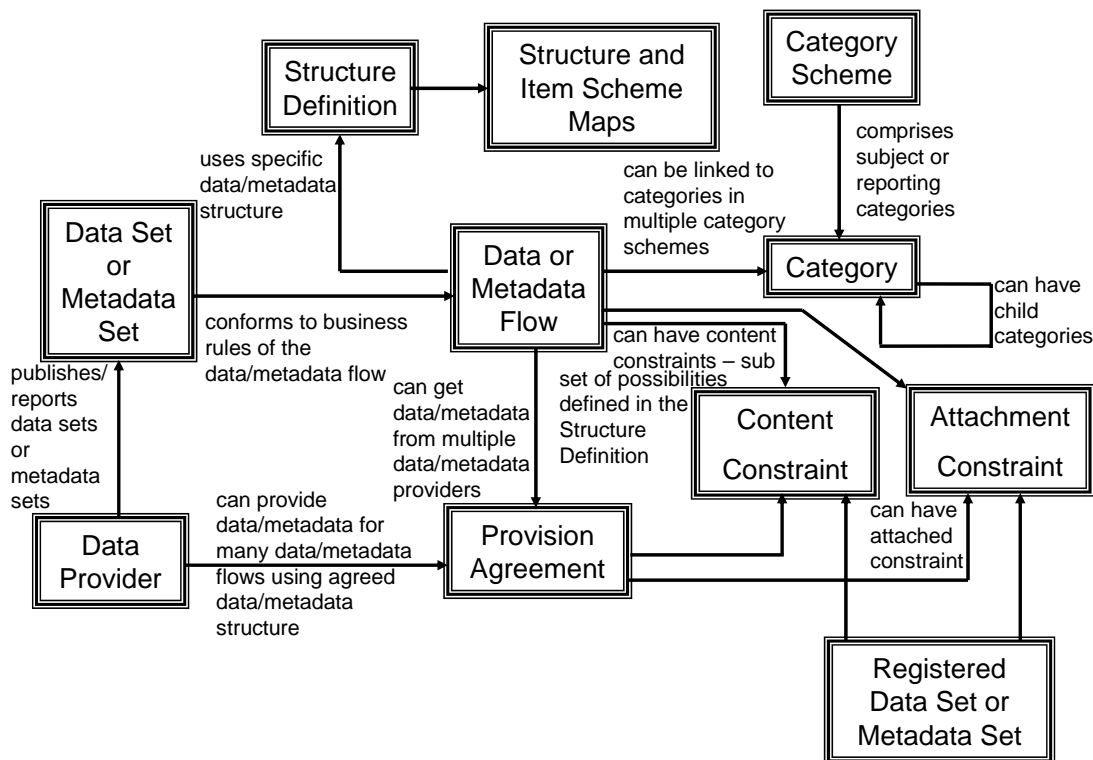

---
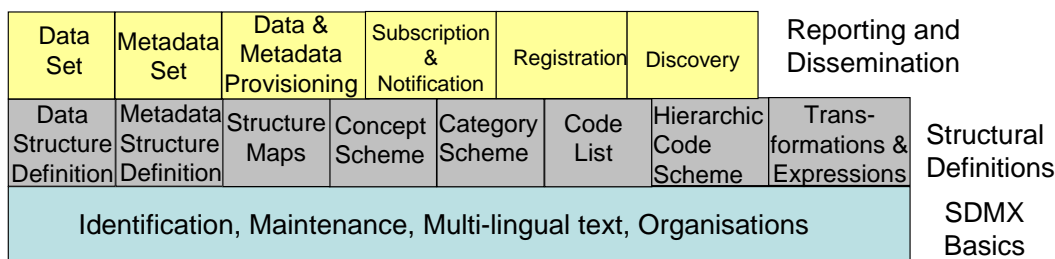
[1] Reference Metadata: Metadata describing the contents and the quality of the statistical data (MCV Preliminary Draft March 2006)

SDMX: the Big Picture

Scope

In order to show how SDMX supports this it is important to see SDMX from a high level perspective.

The scope of the SDMX standards can be best described in terms of the diagram below.

This diagram is described below. It is important to understand at the outset that SDMX can be used at any level in the data provision or dissemination chain provided that the data are aggregated, and the reference metadata aspects of SDMX can be used for all metadata regardless of what it is or for which object it relates. So, for a National Statistical Institute (NSI), SDMX can be used for upwards reporting at the European or International level, or the NSI can use SDMX for the reporting of data and metadata from its constituent organisations. Similarly, an organisation can use SDMX to define the structures that support data and metadata dissemination.

Explanation of the Big Picture

Data Sets and Metadata Sets are reported over time by Data Providers according to certain business rules. These organisations can operate at the national level reporting to Community or International Organisations, at the sub national level reporting to the national level, or at any level to support dissemination. The business rules include the agreed format of the data and metadata and the reporting schedule. In SDMX terms these business rules relate to the Data Flow or Metadata Flow. The Data Flow and Metadata Flow describe the provision of data and metadata over time. They are analogous to the concept of a periodical in magazine publishing: the periodical describes the artefact for which a magazine is published on, say, a

monthly basis – at the very least the data held for the periodical would be its publication dates, but would also link to other information such as distribution channels, advertising sponsors etc. The Data Set or Metadata Set is analogous to the specific magazine (e.g. the June 2007 edition). For data and metadata the Data Set/Metadata Set contains the actual data and metadata values for a specific period or periods, whereas the Data Flow/Metadata Flow holds or links to the other information. Importantly, the Data Flow/Metadata Flow can be linked to one or more subject matter domains (Category), and must be linked to exactly one Structure Definition (Data Structure Definition or Metadata Structure Definition). The subject matter domain link allows for a drill down for queries for data or to assemble data sets that are similarly structured.

The Structure Definitions includes Concepts and Code Lists and their link to the Data Structure and Metadata Structure Definitions that use them.

In any reporting scenario (e.g. from national level to international level, or from sub national to national level) one Data Provider can report data or metadata for many Data of Metadata Flows, and for one Data or Metadata Flow there may be many Data Providers. The Provision Agreement is the union of these two – it contains the details of the reporting of data or metadata by a specific Data Provider for a specific Data or Metadata Flow. Note that in the SDMX model the term Data Provider refers to a provider of data or (reference) metadata.

A Constraint may be applied to either the Data or Metadata Flow or the Provision Agreement. The purpose of the Constraint is to define the sub set of the entire universe of data or metadata implied by the Data or Metadata Structure Definition that is applicable to either the Data or Metadata Flow or the provision of data for that Data or Metadata Flow by a specific Data Provider (i.e. for the Provision Agreement). The Constraint is also used in a registry centric scenario to hold the details of the registered Data or Metadata Set.

This paper is confined to the use of SDMX for defining metadata structures and reporting metadata sets. This is achieved by means of an example.

Example

The Metadata

The example is taken from the Eurostat website – the Summary Methodology for the National Accounts. Some snapshots of these metadata are shown below.

Quarterly National Accounts
Eurostat Metadata in SDDS format: Summary Methodology

GEOGRAPHICAL AREA	European Union (EU-25), euro area, EU Member States, Candidate Countries, Norway, Iceland, Switzerland, US, Japan
DATA CATEGORY	Quarterly National accounts: main aggregates, breakdowns and auxiliary indicators

Last update of this document: 30 November 2006

Concepts, definitions and classifications
Scope / coverage of the data
Accounting conventions
Nature of the basic data
Compilation practices
Other aspects

Base Page

Contact

Eurostat, Statistical Office of the European Communities
Unit C2: National accounts - Production
L-2920 Luxembourg
For any question on data and metadata, please contact:
EUROPEAN STATISTICAL DATA SUPPORT

2. Scope / coverage of the data

GEOGRAPHICAL COVERAGE

Eurostat computes aggregates for the EU and the euro area as a whole. Moreover, when it is possible, it publishes data on Quarterly National Accounts aggregates for Member States separately, EFTA countries and Candidate Countries, the United States and Japan. Since countries which are neither Member State, Candidate Country of EFTA Member have no legal obligation to transmit data to Eurostat, coverage is usually limited to key aggregates, and delays may be longer.

The length of the series available varies widely from one country to another, partly due to derogations provided for in the transmission and back-projection programme. The default minimum period covered is currently from 1995 onwards, with some exceptions in the case of the New Member States. For many countries, however, longer series are available.

The allocation of FISIM to user sectors has not been applied yet by all Member States to the whole length of time series previously covered. This can result in breaks in the time series shown.

Since National figures are collected from National Statistical Institutes' Accounts Departments, for further information you may also refer to National Statistical Institutes and National Central Banks.

STATISTICAL UNITS AND STATISTICAL POPULATION

National accounts are dealing with the economy (or large sub-sectors) as a whole. It combines data from a host of base statistics, and thus has no common sampling reference frame. The elementary building block of ESA95 statistics is the institutional unit (ESA95, 2.12.), "an elementary economic decision-making centre characterised by uniformity of behaviour and decision-making autonomy in the exercise of its principal function". This can be, amongst others, a household, a corporation or a government agency. Institutional units producing goods and services are often engaged in a combination of activities at the same time. For national accounts purposes, they are therefore split into local kind-of-activity units (ESA 1995, 2.102.), characterised by involvement in a single activity. These are then grouped into industries, so that a big industrial enterprise may contribute to activities in a number of different branches. For further detail, please refer to ESA95.

3. Accounting conventions

REFERENCE PERIOD

The periods referred to are quarters, with temporal coverage varying between geographical units.

BASE PERIOD

The base year for computation of constant prices is traditionally a single, fixed year, which is moved ahead about each five years. The whole time series available is then expressed in prices of the new base year. This practice has the unpleasant feature that the further one moves away from the base year, the more irrelevant becomes the price structure of the base year for the economic reality. In particular for economic activities in dynamic fields with rapidly moving prices (such as information and communication technologies) expressing growth in prices of a distant year leads to serious distortions. This is why Commission Decision 98/715/EC demands that the base year shall be the previous year. This guarantees that volumes are measured using the most recent price structure. However, this also means that the base is moved ahead depending on the observation period, and no two years have the same price base, so that volume growth rates cannot be calculated directly from series at previous year's prices. Multiplying successive growth rates starting from an arbitrary reference year's level will give a true volume time series. Due to its construction, this is called a chain-linked series.

Defining an MSD

The Tasks

The tasks that need to be undertaken in defining the metadata structure definition (MSD) are:

1. Analysis of the entire set of metadata in order to identify and document the "Concepts" for which metadata are to be reported or disseminated.

2. Determine the structure of the "Metadata Report" in terms of the concepts used, the hierarchy of the concepts when used in the report, and their "representation" (e.g. is a code list used, is the format free text?).

3. Specify the "object type" to which the metadata are to be attached, and how this object type is identified: knowledge of the SDMX Information model is useful here, as the metadata can only be attached to object types that can be identified in terms of the object types that exist in the information model.

It does not really matter in which order these tasks are performed. In this example we will start with structure of the metadata report (task 2), and from this derive the Concepts (task 1), and then identify where the metadata are to be attached (task 3).

Metadata Report Structure

The following screenshots show some of the metadata reported by Eurostat on the Eurostat statistical web site.

Contact Information

Contact

Eurostat, Statistical Office of the European Communities
Unit C2 – National accounts - Production
L-2920 Luxembourg
For any question on data and metadata, please contact:
EUROPEAN STATISTICAL DATA SUPPORT

In this case, there is no individual name, just the organisation and the organisation unit. Also, there is no telephone number or fax number, but there is a web contact address (clicking on "EUROPEAN STATISTICAL DATA SUPPORT" reveals http://epp.eurostat.cec.eu.int/pls/portal/url/page/PGP_DS_SUPPORT

From this information the following report structure, and underlying concepts, can be derived.

In this table there are two levels of hierarchy in the report – called "Attribute" and "Sub Attribute". The actual definition of the concept is in the Concept Scheme. The usage of the concept, its place in the hierarchy, representation, and attachment are defined in the "Metadata Attribute" of the MSD (called Attribute in the tables below).

Attribute	Concept	Sub Attribute	Concept	Format
Contact	CONTACT	Contact Organization Name	ORGANIZATION_ NAME	text
		Contact Organization Postal Zone	POSTAL_ZONE	text
		Contact Organization Unit	ORGANIZATION_ UNIT	text
		Contact web Address	WEB _ADDRESS	text

Content Metadata

2. Scope / coverage of the data

GEOGRAPHICAL COVERAGE

Eurostat publishes national accounts data for European Union (EU25), euro area, EU Member States, EFTA countries, Candidate Countries, the United States and Japan. Eurostat estimates the aggregates for EU25 and euro area (as well as for the EU15 to the extent this geographical aggregate is supported); all the other data are produced by the statistical offices of the respective countries. For further information on country data you may also refer to National Statistical Institutes and National Central Banks (links given on the Eurostat web site).

Member States, ETFA countries and Candidate Countries have legal obligations to submit their data to Eurostat. These data are the inputs for Eurostat's estimates of EU25 and euro area. Since the United States and Japan have no such obligation to transmit data to Eurostat, coverage is usually limited to key aggregates, and delays between national publication and availability on Eurostat's website may be longer. Note that for some countries not covered in the national accounts domain, some central aggregates such as GDP may be available from the general statistics domain of Eurostat's online database.

STATISTICAL UNITS

National accounts are dealing with the economy (or large sub-sectors) as a whole. It combines data from a host of base statistics, and thus has no common sampling reference frame. The elementary building block of ESA95 statistics is the institutional unit (ESA95, 2.12.), "an elementary economic decision-making centre characterised by uniformity of behaviour and decision-making autonomy in the exercise of its principal function". This can be, amongst others, a household, a corporation or a government agency. Institutional units producing goods and services are often engaged in a combination of activities at the same time. For national accounts purposes, they are therefore split into local kind-of-activity units (ESA 1995, 2.102.), characterised by involvement in a single activity. These are then grouped into industries, so that a big industrial enterprise may contribute to activities in a number of different branches. For further detail, please refer to ESA95.

3. Accounting conventions

REFERENCE PERIOD

The accounting period is the calendar year, with temporal coverage varying between geographical units.

BASE PERIOD

When flows and stocks are valued at the price level in the accounting period they are said to be valued at current prices. Valuation at constant prices (ESA95, 1.25d, 1.56) means valuing flows and stocks at the price of a previous period (called base year). The purpose of the valuation at constant prices is to assess the dynamics of economic development irrespective of price movements (see also Eurostat's "Handbook of price and volume measurement" for more details). This is achieved by decomposing changes of values over time into changes in prices and changes in volume. Price, value and volume are related via the following central equation:

$$\text{Value} = \text{Volume} \times \text{Price}$$

The following attributes can be derived from these examples.

Attribute	Concept Used	Sub Attribute	Concept Used	Format
Scope and Coverage	SCOPE_ COVERAGE	Reference Area	REFERENCE_AREA	text
		Statistical Unit	STATISTICAL_UNIT	text
Accounting Conventions	ACCOUNTING_ CONVENTIONS	Reference Period	REFERENCE_PERIOD	
		Base Period	BASE_PERIOD	

Concept Scheme

The following concepts are derived from the tables above.

CONTACT
ORGANISATION_NAME
POSTAL_ZONE
ORGANISATION_UNIT
WEB_ADDRESS
SCOPE_COVERAGE
REFERENCE_AREA
STATISTICAL_UNIT
ACCOUNTING_CONVENTIONS
REFERENCE_PERIOD
BASE_PERIOD

The concepts in the concept scheme can be defined in a hierarchy where there is a semantic link between the parent and child concepts; the child concept(s) having a more fine grained semantic meaning of (a part of) the parent.

The Attachment Object Type

The Metadata Set which is reported (i.e. the actual metadata content) is intended to be metadata about "something". The "something" is the object type and in an MSD it is necessary to declare the object type and to define how it is identified in terms of its constituent components: for instance, a Code would be identified by a combination of the Code List identifier and the Code identifier.

The attachment object type must be definable using the identifiable object types in the SDMX Information Model – the XML schema demands this and list the following object types.

```
Agency
ConceptScheme
Concept
Codelist
Code
KeyFamily
Component
KeyDescriptor
MeasureDescriptor
AttributeDescriptor
GroupKeyDescriptor
Dimension
Measure
Attribute
CategoryScheme
ReportingTaxonomy
Category
OrganisationScheme
DataProvider
MetadataStructure
FullTargetIdentifier
PartialTargetIdentifier
MetadataAttribute
DataFlow
ProvisionAgreement
MetadataFlow
ContentConstraint
AttachmentConstraint
DataSet
XSDataSet
MetadataSet
HierarchicalCodelist
Hierarchy
StructureSet
StructureMap
ComponentMap
CodelistMap
CodeMap
CategorySchemeMap
CategoryMap
OrganisationSchemeMap
OrganisationRoleMap
ConceptSchemeMap
ConceptMap
Process
ProcessStep
```

The object type for this example is indicated in the following extract from the screen shot of the website.

Quarterly National Accounts
Eurostat Metadata in SDDS format: Summary Methodology

GEOGRAPHICAL AREA	European Union (EU-25), euro area, EU Member States, Candidate Countries, Norway, Iceland, Switzerland, US, Japan
DATA CATEGORY	Quarterly National accounts: main aggregates, breakdowns and auxiliary indicators

Last update of this document: 30 November 2006

The object type is the Data Category (called "Category" in the SDMX Information Model). If the intent of the MSD is to define where the metadata are to be attached in the Eurostat dissemination environment then this is all that is required. However, if Eurostat wish to publish this and make it available to other organisations (e.g. as a downloadable file) then it would be necessary to also identify the Data Provider (which in this case is Eurostat). Both these object types could be associated with a coding scheme – there would certainly be a list for all of the data categories (this would be a "Category Scheme"), but for the Data Provider this could be declared as non enumerated (i.e. text).

The resultant MSD could therefore be used to declare metadata (in a metadata set) for any data category present in the data category scheme.
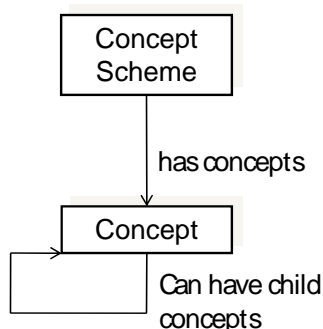
Bringing it Together

Concept Scheme

A concept scheme is a "container" for concepts. In SDMX this is the level at which the concepts are maintained. The concept scheme has a maintenance agency, identity and versioning information.

Concepts in a concept scheme can be hierarchic. This hierarchy represent a semantic hierarchy where child concepts are semantically linked to the parent concept, and typically each child would have a semantic that is a part of the parent. Often concepts are used in reporting hierarchies and these hierarchies are built in the Report Structure of the Metadata Structure Definition.

An example of a concept scheme would be the DQAF scheme of the IMF. Other organisations will have concept schemes and SDMX allows the concepts in one scheme to be related to concepts in other concept schemes.

A schematic of the structure of a concept scheme is shown below.

```
           ┌──────────┐
           │ Concept  │
           │ Scheme   │
           └────┬─────┘
                │
           has concepts
                │
                ▼
           ┌──────────┐
        ┌─►│ Concept  │
        │  └────┬─────┘
        └───────┘  Can have child
                   concepts
```

An extract of the SDMX-ML for the concept scheme is shown below.

```
<structure:ConceptScheme id="SDMX_CONCEPTS" agencyID="EUROSTAT">
<structure:Name>DQAF Concept Scheme</structure:Name>
<structure:Concept id="SCOPE_COVERAGE">
            <structure:Name xml:lang="en">Scope and Coverage</structure:Name>
            <structure:Description xml:lang="en">The coverage of the statistics and
how consistent this is with internationally accepted standards, guidelines, or good
practices. The scope/coverage includes a description of target population, and
geographic, sector, institutional, item, population, product, and other
coverage.</structure:Description>
</structure:Concept>
<structure:Concept id="REFERENCE_AREA">
            <structure:Name xml:lang="en">Reference Area</structure:Name>
            <structure:Description xml:lang="en">A boundary or constraint expressed
in terms of geographical, political, terriorial or other method of
division.</structure:Description>
</structure:Concept>
<structure:Concept id="STATISTICAL_UNIT">
            <structure:Name xml:lang="en">Statistical unit/reporting
unit</structure:Name>
            <structure:Description xml:lang="en">
Statistical unit = an object of statistical survey and the bearer of statistical
characteristics. The statistical unit is the basic unit of statistical observation
within a statistical survey.</structure:Description>
</structure:Concept>
<structure:Concept id="ACCOUNTING_CONVENTIONS">
            <structure:Name xml:lang="en">Accounting Conventions</structure:Name>
            <structure:Description xml:lang="en">The practical aspects and
conventions used when compiling data from diverse sources under a common methodological
framework. It may refer to descriptions of the types of prices used to value flows and
stocks, or other units of measurements used for recording the phenomena being observed;
the time of recording of the flows and stocks or the time of recording of other
phenomena that are measured; and the grossing/netting procedures that are
used.</structure:Description>
</structure:Concept>
```
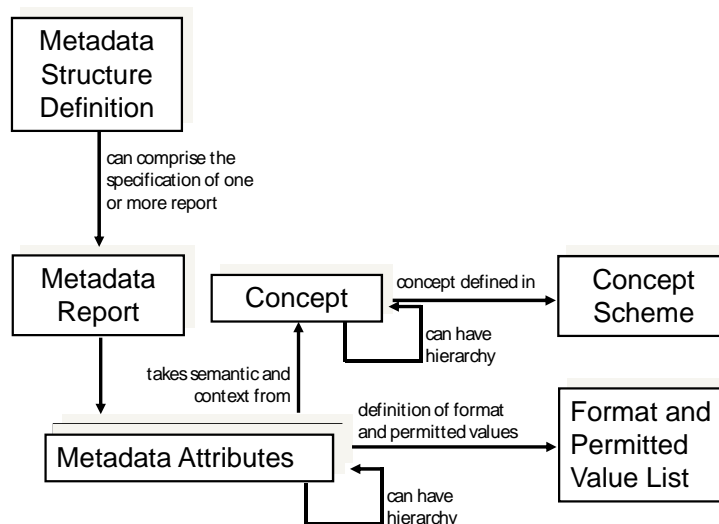
Report Structure

General Structure

The Report Structure is defined within a Metadata Structure Definition.


The Metadata Report is given an Id. and name and the Metadata Attributes that comprise the report are defined. Each attribute must reference a Concept and the Concept Scheme within which it is maintained. Thus the Metadata Attributes can be based on concepts from different

concept schemes. The format or permitted value list (e.g. a code list, and whether its presence in a Metadata Set is mandatory) is declared for the Metadata Attribute.

Contact Report

For the Contact report the following schematic shows the definition.
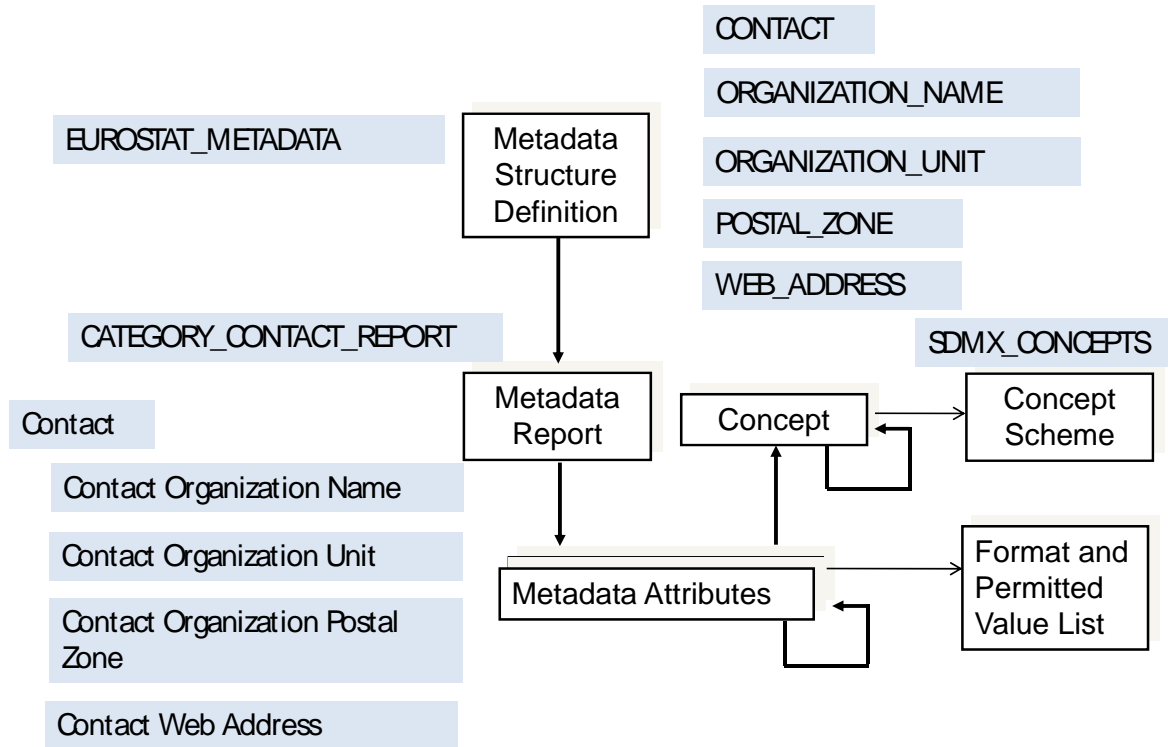
In this example the Contact Organization Name, Contact Organization Unit, and the Contact Organization Postal Zone are child Metadata Attributes of the Contact, even though the underlying Concepts are not hierarchically linked. Thus the report structure allows the definition of a reporting hierarchy even though there is semantic hierarchy declared in the Concept Scheme.

An extract of the SDMX-ML for this report structure is shown below.

```
<structure:ReportStructure id="CATEGORY_CONTACT_REPORT" target="CATEGORY">
      <structure:Name>Category Contact Report</structure:Name>
      <structure:Description>Metadata report structure for contact details reported at the
level of the Data Category</structure:Description>
<structure:MetadataAttribute conceptRef="CONTACT" conceptSchemeRef="SDMX_CONCEPTS"
usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
<structure:MetadataAttribute conceptRef="ORGANIZATION_NAME" conceptSchemeRef="SDMX_CONCEPTS"
usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
            <structure:TextFormat textType="String"/>
            <structure:Annotations>
            <common:Annotation>
            <common:AnnotationType>ATTRIBUTE_NAME</common:AnnotationType>
            <common:AnnotationText>Contact Organisation Name</common:AnnotationText>
            </common:Annotation>
            </structure:Annotations>
      </structure:MetadataAttribute>
<structure:MetadataAttribute conceptRef="ORGANIZATION_UNIT" conceptSchemeRef="SDMX_CONCEPTS"
usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
            <structure:TextFormat textType="String"/>
            <structure:Annotations>
            <common:Annotation>
            <common:AnnotationType>ATTRIBUTE_NAME</common:AnnotationType>
            <common:AnnotationText>Contact Organisation Unit</common:AnnotationText>
            </common:Annotation>
            </structure:Annotations>
      </structure:MetadataAttribute>
<structure:MetadataAttribute conceptRef="POSTAL_ZONE" conceptSchemeRef="SDMX_CONCEPTS"
usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
            <structure:TextFormat textType="String"/>
            <structure:Annotations>
            <common:Annotation>
            <common:AnnotationType>ATTRIBUTE_NAME</common:AnnotationType>
            <common:AnnotationText>Contact Organisation Postal Zone</common:AnnotationText>
            </common:Annotation>
            </structure:Annotations>
      </structure:MetadataAttribute>
<structure:MetadataAttribute conceptRef="WEB_ADDRESS" conceptSchemeRef="SDMX_CONCEPTS"
usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
            <structure:TextFormat textType="String"/>
            <structure:Annotations>
            <common:Annotation>
            <common:AnnotationType>ATTRIBUTE_NAME</common:AnnotationType>
            <common:AnnotationText>Contact web address</common:AnnotationText>
            </common:Annotation>
            </structure:Annotations>
      </structure:MetadataAttribute>
</structure:MetadataAttribute>
</structure:ReportStructure>
```

In this example the annotation facility of SDMX-ML has been used to give a name to the Metadata Attribute. The annotation facility is useful for defining report attribute names (which could be used by an application that generated metadata report forms from an MSD) or a cross reference to an internal id for the Metadata Attribute (such as may be used in a database). In the example above the Attribute Name is declared.

Quality Report

The following schematic shows the definition of the quality report.
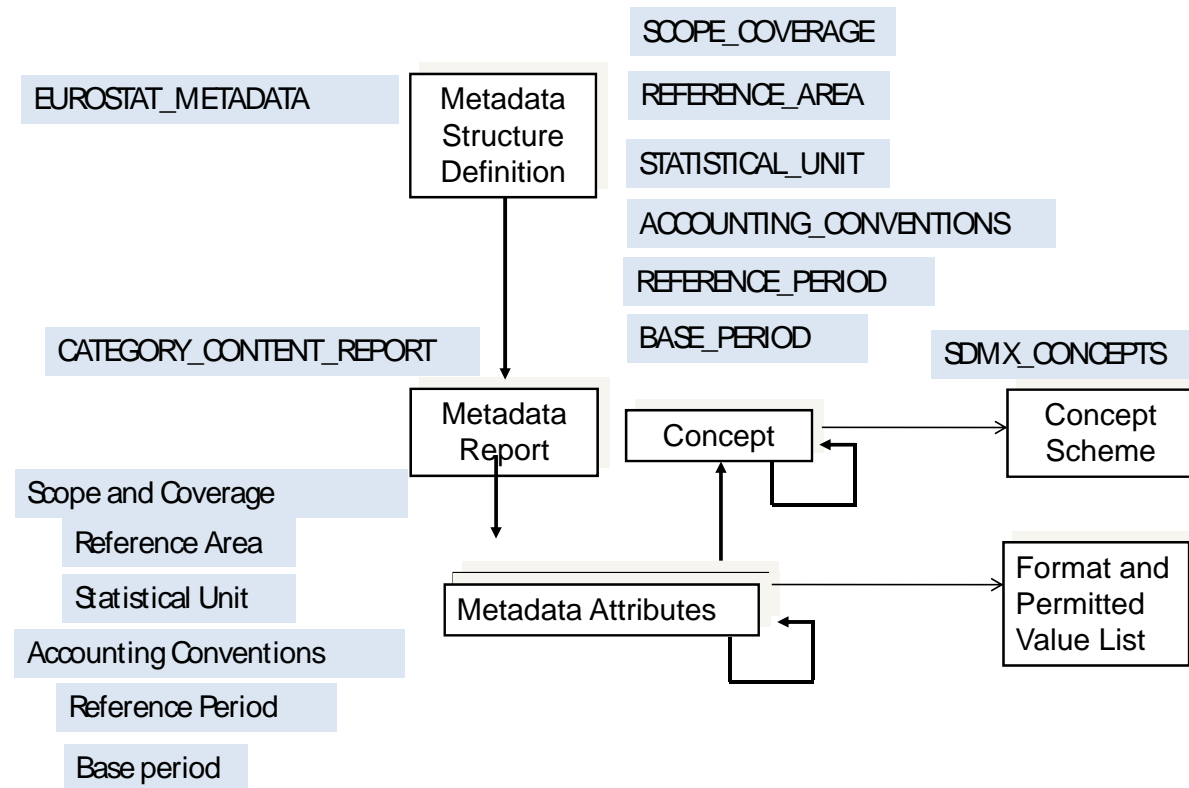
An extract of the SDMX-ML for this report structure is shown below.

```
<structure:ReportStructure id="CATEGORY_CONTENT_REPORT" target="CATEGORY">
       <structure:Name>Category Quality Report</structure:Name>
       <structure:Description>Metadata report structure for quality metadata reported at the
level of the Data Category</structure:Description>
<structure:MetadataAttribute conceptRef="SCOPE_COVERAGE" conceptSchemeRef="SDMX_CONCEPTS"
usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
<structure:MetadataAttribute conceptRef="REFERENCE_AREA" conceptSchemeRef="SDMX_CONCEPTS"
usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
</structure:MetadataAttribute>
<structure:MetadataAttribute conceptRef="STATISTICAL_UNIT" conceptSchemeRef="SDMX_CONCEPTS"
usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
</structure:MetadataAttribute>
</structure:MetadataAttribute>
<structure:MetadataAttribute conceptRef="ACCOUNTING_CONVENTIONS"
conceptSchemeRef="SDMX_CONCEPTS" usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
<structure:MetadataAttribute conceptRef="REFERENCE_PERIOD" conceptSchemeRef="SDMX_CONCEPTS"
usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
</structure:MetadataAttribute>
<structure:MetadataAttribute conceptRef="BASE_PERIOD" conceptSchemeRef="SDMX_CONCEPTS"
usageStatus="Mandatory" conceptSchemeAgency="EUROSTAT">
</structure:MetadataAttribute>
</structure:MetadataAttribute>
```

Note that, for brevity, in this extract the annotations have been omitted.

Defining the Attachment Object Type

Schematic
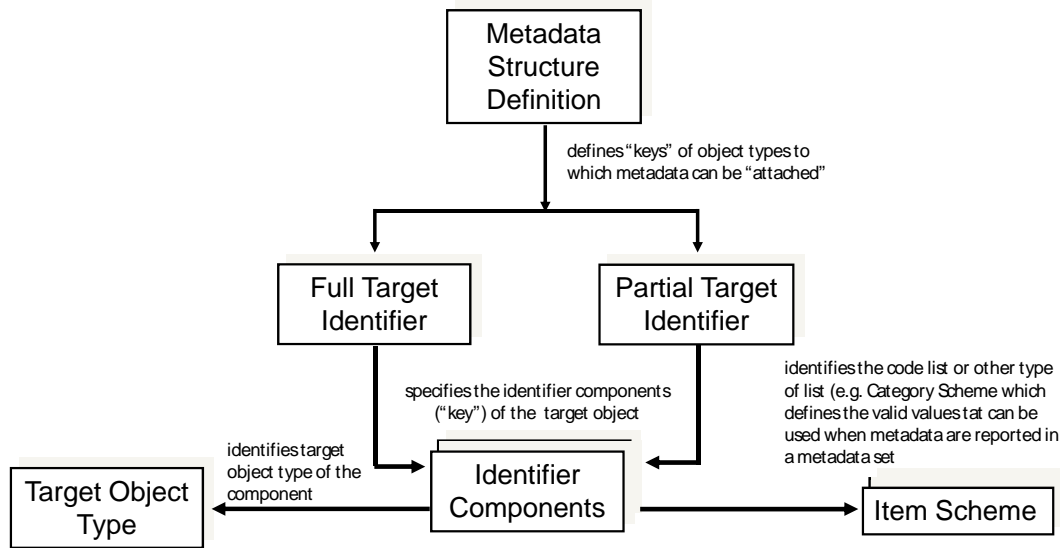
The intent here is to define the object types and identifiers of the object types. These object types (Target Object Type in the diagram) must be in the list in the schema (this list is shown earlier in this paper). There can be many such objects identified in a single Metadata Structure Definition. The Full Target Identifier defines all of the possible object types that are within the scope of the MSD, and links each with a representation scheme (which can be any of the "Item Schemes" in the Information Model e.g. Code List, Concept Scheme, Category Scheme, Organisation Scheme). The Partial Target Identifier references a sub set of the Identifier Components of the Full Target Identifier. As will be seen below, a Report Structure is linked to either the Full Target Identifier or one of the Partial Target Identifier.

Attachment Object Types

In this example the metadata are attached to a Data Category (i.e. a Category in a Category Scheme). The metadata are attached to a number of child categories of the Structural Indicators. Note that the Geographical Area is additional metadata and is not a part of the attachment mechanism.

In addition to the Category it is also necessary to identify the Data Provider (in this case Eurostat), as many organisations may have such metadata and it is important to identify that this is metadata about Eurostat's published data. It is probable that some metadata will be attached to just the Data Provider.
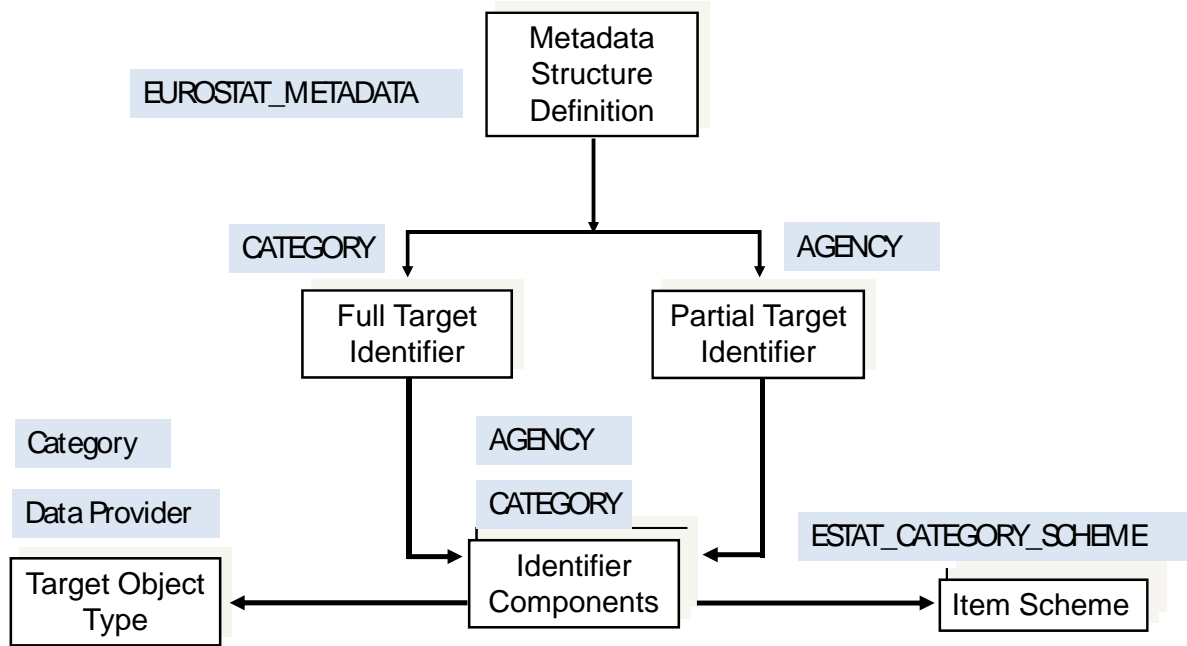
In this MSD, the Full Target Identifier (called CATEGORY) comprises the Category and the DataProvider object types. The Partial Target Identifier (AGENCY) references just the Identifier Component linked to the DataProvider (called AGENCY).

An extract of the SDMX-ML for this section of the MSD is shown below.

```
<MetadataStructureDefinitions>
        <structure:MetadataStructureDefinition id="EUROSTAT_METADATA" agencyID="EUROSTAT">
                <structure:Name xml:lang="en">Content Metadata</structure:Name>
                <structure:TargetIdentifiers>
                <structure:FullTargetIdentifier id="CATEGORY">
                        <structure:Name xml:lang="en">Subject Domain Category</structure:Name>
                        <structure:IdentifierComponent id="AGENCY">
                                <structure:Name xml:lang="en">Agency
Organisation</structure:Name>

        <structure:TargetObjectClass>DataProvider</structure:TargetObjectClass>
                        </structure:IdentifierComponent>
                        <structure:IdentifierComponent id="CATEGORY">
                                <structure:Name xml:lang="en">Subject Domain
Category</structure:Name>

        <structure:TargetObjectClass>Category</structure:TargetObjectClass>
                                <structure:RepresentationScheme
representationScheme="ESTAT_CATEGORY_SCHEME" representationSchemeAgency="EUROSTAT"
representationSchemeType="Category"/>
                        </structure:IdentifierComponent>
        </structure:FullTargetIdentifier>
        <structure:PartialTargetIdentifier id="AGENCY">
                <structure:Name xml:lang="en">Agency</structure:Name>
                <structure:IdentifierComponentRef>AGENCY</structure:IdentifierComponentRef>
                </structure:PartialTargetIdentifier>
                </structure:TargetIdentifiers>
<!—this is followed by the various Report Structures -->
```

Note that this metadata is attached at a fairly high level – the level of the subject domain category – for the data provider. If there are metadata at a lower level of granularity, for instance at the level of the "table" in the list below, then this can also be specified in an MSD.

Table 0101 (Gross value added at basic prices and gross domestic product at market prices, A6 breakdown),
Table 0102 (Gross domestic product from the expenditure side),
Table 0103 (Gross domestic product from the income side),
Table 0104 (Final consumption),
Table 0105 (Gross capital formation),
Table 0106 (Exports and imports of goods (fob) and services),
Table 0107 (Disposable income),
Table 0108 (Saving and net lending/borrowing),
Table 0109 (Real disposable income),
Table 0110 (Population and employment),
Table 0111 (Employment by industry, A6 breakdown),
Table 0112 (Compensation of employees by industry, A6 breakdown),
Table 0113 (Gross value added in A17 breakdown),
Table 0114 (Employment by industry in A17 breakdown),
Table 0115 (Compensation of employees by industry in A17 breakdown),
Table 0116 (Final consumption expenditure of households by purpose (COICOP)),
Table 0118 (Gross operating surplus by industry in A17 breakdown),
Table 0120 (Exports of goods (f.o.b.) and services by Member States of the European Union/third Countries,
Table 0121 (Imports of goods (f.o.b.) and services by Member States of the European Union/third Countries.

In order to attach metadata to each of the tables then each of these tables can be defined as a "Dataflow" and the metadata attached to the provision of the data by a data provider for this dataflow.

It is also possible, using an MSD, to define metadata attachment to data structure definitions, components of data structure definitions (e.g. a dimension), even values of dimensions.

The final piece of the jigsaw in this example is to link the Report Structures to the relevant Full or Partial Target Identifier. This has already been done in the XML examples, the schematic is shown below.
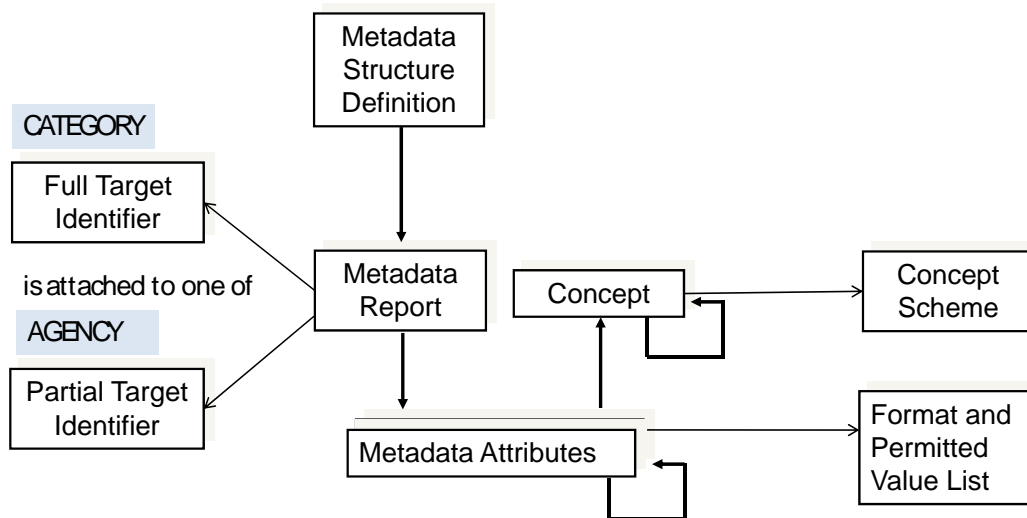
The XML that makes this link is the target attribute in the Report Structure as shown in the example below.

```
<structure:ReportStructure id="CATEGORY_QUALITY_REPORT" target="CATEGORY">
        <structure:Name>Category Quality Report</structure:Name>
```

Metadata Set

General Schematic

Metadata is reported in a Metadata Set. Each metadata report is reported in a separate Metadata Set, and there can be many such sets in an SDMX Message. The schematic of the structure of a metadata set is shown below.
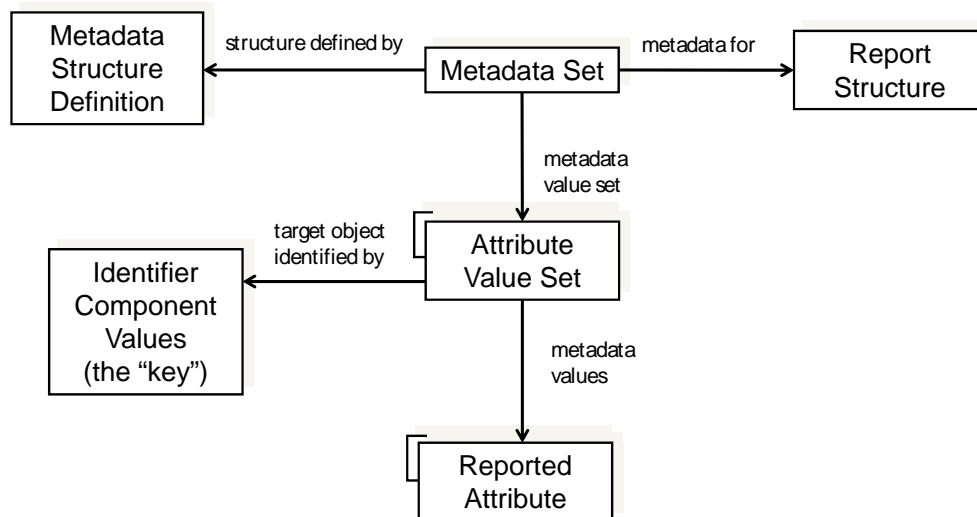
The Contact Metadata Set

Contact

Eurostat, Statistical Office of the European Communities
Unit C2 – National accounts - Production
L-2920 Luxembourg
For any question on data and metadata, please contact:
EUROPEAN STATISTICAL DATA SUPPORT

A schematic of the Metadata Set for the Contact Report is shown below.

EUROSTAT_METADATA

```
Metadata
Structure
Definition
```

CATEGORY_CONTACT_REPORT

```
Report
Structure
```

Metadata Set

Category=Key_Indicators.Structural_Indicators

Data Provider=EUROSTAT

```
Identifier
Component
Values
(the "key")
```

```
Attribute
Value Set
```

CONTACT

ORGANISATION_UNIT=Unit C2 National accounts : production

ORGANISATION_NAME=Eurostat, Statistical Office of the European Communities

POSTAL_ZONE=L-2920 Luxembourg

WEB_ADDRESS= http://....

```
Reported
Attribute
```

Contact
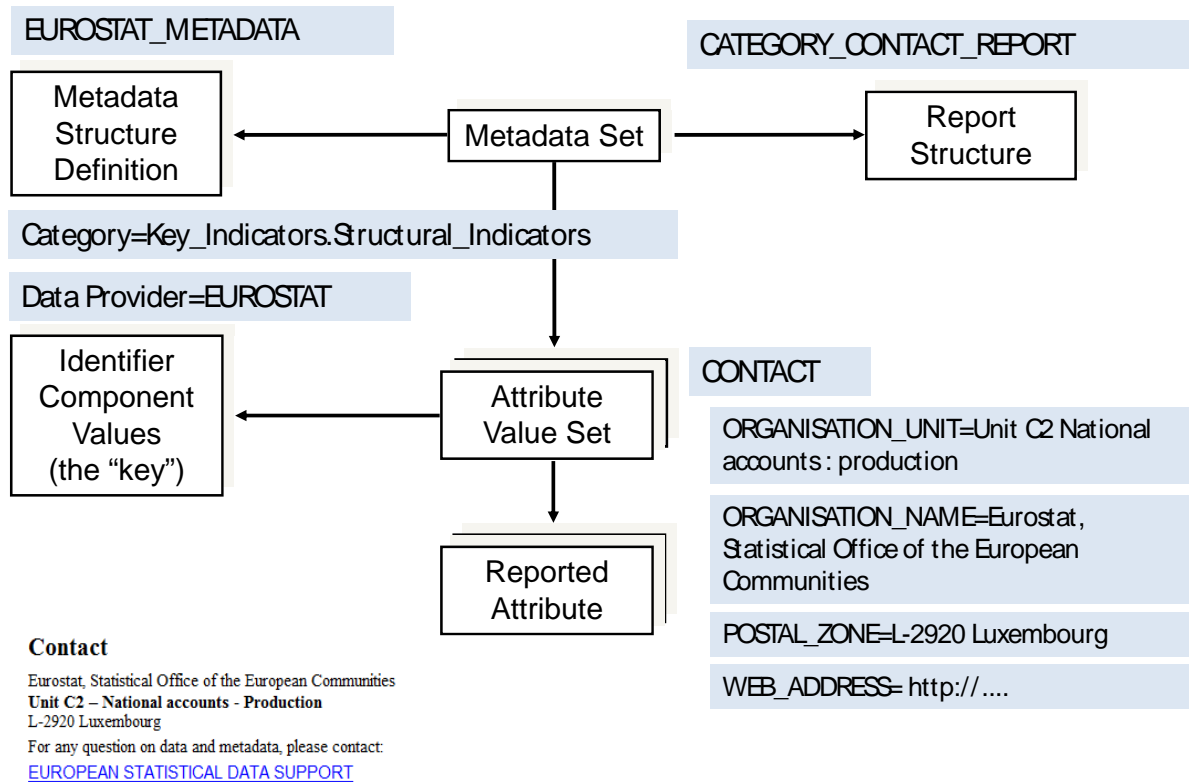Eurostat, Statistical Office of the European Communities
Unit C2 – National accounts - Production
L-2920 Luxembourg
For any question on data and metadata, please contact:
EUROPEAN STATISTICAL DATA SUPPORT

The SDMX-ML for this, reported in the generic format for a Metadata Set is:

```xml
<genericmetadata:MetadataSet>
<genericmetadata:MetadataStructureRef>EUROSTAT_METADATA</genericmetadata:MetadataStructureRef>
<genericmetadata:MetadataStructureAgencyRef>EUROSTAT</genericmetadata:MetadataStructureAgencyRef>
<genericmetadata:ReportRef>CATEGORY_CONTACT_REPORT</genericmetadata:ReportRef>
        <genericmetadata:AttributeValueSet>
        <genericmetadata:TargetRef>CATEGORY</genericmetadata:TargetRef>
        <genericmetadata:TargetValues>
        <genericmetadata:ComponentValue
object="Category">Structural_Indicators</genericmetadata:ComponentValue>
        <genericmetadata:ComponentValue
object="DataProvider">Eurostat</genericmetadata:ComponentValue>
        </genericmetadata:TargetValues>
        <genericmetadata:ReportedAttribute conceptID="CONTACT">
        <genericmetadata:ReportedAttribute conceptID="ORGANIZATION_UNIT">
        <genericmetadata:Value>Unit C2 National accounts : production</genericmetadata:Value>
        <genericmetadata:ReportedAttribute conceptID="ORGANIZATION_UNIT">
        <genericmetadata:Value>Unit C5 Validation of public accounts</genericmetadata:Value>
        </genericmetadata:ReportedAttribute>
        <genericmetadata:ReportedAttribute conceptID="ORGANIZATION_NAME">
        <genericmetadata:Value>Eurostat, Statistical Office of the European
Communities</genericmetadata:Value>
        </genericmetadata:ReportedAttribute>
        <genericmetadata:ReportedAttribute conceptID="POSTAL_ZONE">
        <genericmetadata:Value>L-2920 Luxembourg</genericmetadata:Value>
        </genericmetadata:ReportedAttribute>
        <genericmetadata:Value>http://epp.eurostat.cec.eu.int/pls/portal/url/page/PGP_DS_SUPPORT</g
enericmetadata:Value>
        </genericmetadata:ReportedAttribute>
</genericmetadata:ReportedAttribute>
</genericmetadata:AttributeValueSet>
</genericmetadata:MetadataSet>
```

Metadata is Everywhere

Of course, metadata can be applied to any object at any time and the SDMX standards recognise this fact and support it. The examples show that users can define any metadata reporting scheme required (in terms of concepts and the valid values that the concept can have in a metadata set) for any object that can be identified.

The structural mechanism used is similar to the definition of attributes in a data structure definition, but whereas the metadata structure definition can define metadata for any type of object, the object of concern in a data structure definition is limited to:

- A data set
- A full series key – the attribute applies to the key and therefore all observation values for that key in the specific data set
- A partial key that, in effect, identifies a slice of the data observations
- A specific observation – identified by the full key and the time period

An important point to note is that the mechanisms used for the definition of metadata structures and data structures are similar in nature: they are what are termed "metamodels". The metamodelling approach is one that is necessary for both data and reference metadata because there is a need to describe any type of data and metadata. This results in a very flexible mechanism but its utility for data and metadata inter-operability is dependent upon communities agreeing common concepts and code lists, and in the SDMX community this work is carried out within the group responsible for the content oriented guidelines.