SC8DI07171

D02.01.01.02: StatDCAT-AP – DCAT Application Profile for description of statistical datasets, Draft 2

Document Metadata

Date	2016-05-10
Status	Internal draft
Version	0.05
Authors	Makx Dekkers – AMI Consult Chris Nelson – Metadata Technologies
Reviewed by	Nikolaos Loutas – PwC EU Services Marco Pellegrino – Eurostat Norbert Hohn – Publications Office
Approved by	

This report was prepared for the ISA Programme by:

PwC EU Services

Disclaimer:

Document History

Version	Date	Description	Action
0.01	2016-04-20	Creation of new draft	Creation
0.02	2016-04-29	Addition of proposed extensions in section 6.8 and mapping SDMX to DCAT in section 9	Update
0.03	2016-05-05	Updated sections 1.4, 3.2.3, 3.2.4, 5.2, 5.3, 5.4, 6.1, 6.2, 7 and 8; added sections 6.7.7 and 6.7.8, 9 and 10; made minor edits throughout.	Update
0.04	2016-05-10	Updated sections 9 and 10, added Annex I	Update
0.05	2016-05-10	Changed wording in sections 6.7.1 and 6.7.2	Update

Table of Contents

1 Introduction 1
1.1 Background 1
1.2 Objectives 2
1.3 Roadmap 3
1.4 Structure of this document 4
2 Terminology used in this document 5
3 Related work 6
3.1 Statistical data and metadata initiatives 6
3.1.1 Eurostat and EU Publications Office collaboration 6
3.1.2 SDMX 6
3.1.3 ESMS 8
3.2 Open Data standards and application profiles 8
3.2.1 W3C DCAT 8
3.2.2 DCAT-AP for open data portals in Europe 9
3.2.3 GeoDCAT-AP 11
3.2.4 The Data Cube Vocabulary 12
4 Use cases 13
4.1 Improve discoverability of statistical datasets on open data portals 13
4.2 Federation of open data portals 13
5 Methodology 14
5.1 ISA Core Vocabulary process and methodology 14
5.2 Analysis and decision framework 14
5.3 Stakeholders 14
5.4 Time plan 15
6 The StatDCAT-AP data model 16
6.1 Informal description 16
6.2 Extensions for description of statistical datasets 17
6.3 Overview of the model 17
6.4 Namespaces 18
6.5 UML Class diagram 18
6.6 Description of classes 18
6.6.1 Mandatory Classes 18
6.6.2 Recommended Classes 18

6.6.3 Optional Classes 19
6.7 Description of properties per class 20
6.7.1 Proposed extension: Number of observations 20
6.7.2 Proposed extension: Number of data series 20
6.7.3 Proposed extension: Link to visualisation 21
6.7.4 Proposed extension: Dimensions as property 21
6.7.5 Proposed extension: Dimensions as keywords 21
6.7.6 Proposed extension: Quality aspects 22
6.7.7 Statistical unit 22
6.7.8 Statistical population 22
6.8 Controlled vocabularies 22
7 Conformance statement 23
7.1 Provider requirements 23
7.2 Receiver requirements 23
8 Mapping and Extraction approaches 24
9 Mapping SDMX to DCAT 25
9.1 Scope 25
9.2 Diagrams 25
9.3 Example 27
9.3.1 Introduction 27
9.3.2 SDMX Annotations 28
9.3.3 Data Catalogue 30
9.3.4 Linking to Categories using Categorisations 30
9.3.5 Dataset 31
9.3.6 Distribution 32
9.3.7 Agent 33
9.4 Summary 33
10 SDMX-based Transformation Mechanism 34
10.1 Scope of this section 34
10.2 Transformation mechanism 34
10.3 Transformation input formats 35
10.3.1 Choice of mechanisms 35
10.3.2 SDMX Structural Metadata 35
10.3.3 SDMX Metadata Set 37

10.4 Advantages and disadvantages of the two transformation formats 41
10.4.1 SDMX Structure Message 41
10.4.2 SDMX Metadata Set 42
10.5 Summary 42
11 Acknowledgements 44
Annex I SDMX Files used for the examples 45
I.1 SDMX Structural Metadata 45
I.2 SDMX Metadata Set 50
I.2.1 Content 50

List of Figures

Figure 1: SDMX Main Components 7
Figure 2: SDMX Information Model: Schematic View 7
Figure 3: DCAT schematic data model 9
Figure 4: DCAT-AP Data Model 10
Figure 5: Data Cube vocabulary overview of key terms and relationships 12
Figure 6: Schematic map of SDMX Classes to DCAT-AP 25
Figure 7: DCAT-AP Model mapped to SDMX Model Classes 27
Figure 8: Metadata Used in the Example Mapping 27
Figure 9: SDMX XML schema specification for Annotation 29
Figure 10: SDMX-DCAT mapping example for the DCAT Catalogue 30
Figure 11: Schematic showing linking of SDMX Categories to other SDMX objects ... 30
Figure 12: Linking Catalogue to DCAT Datasets and Category (Topic) Scheme 31
Figure 13: SDMX-DCAT mapping example for the DCAT Dataset 31
Figure 14: Linking a Dataflow to the SDMX Category (Topic) 32
Figure 15: Linking a Distribution to the SDMX Provision Agreement 32
Figure 16: Linking a Distribution (accessURL) to the SDMX Provision Agreement 32
Figure 17:Linking an Agent to the SDMX Agency 33
Figure 18: Diagram of the flow of metadata though the Intermediary Mechanism 35
Figure 19: From Section 9 - Linking a Distribution (accessURL) to the SDMX Provision Agreement 36
Figure 20: Transformation format - Linking a Distribution (accessURL) to the SDMX Provision Agreement 36
Figure 21: Example Provision Agreement for DCAT-AP Distribution 37

Figure 22: Schematic diagram of the SDMX Metadata Structure Definition 38
Figure 23: Schematic diagram of the SDMX Metadata Set..... 38
Figure 24: Metadata Attributes in the DCAT-AP MSD 39
Figure 25: Example Metadata Attribute Specification..... 39
Figure 26: SDMX catalogue metadata pertaining to the DCAT-AP Catalogue..... 40
Figure 27: SDMX category scheme metadata pertaining to the DCAT-AP Catalogue . 40
Figure 28: SDMX dataset metadata pertaining to the DCAT-AP Catalogue..... 41
Figure 29: SDMX distribution metadata pertaining to the DCAT-AP Catalogue 41

1 INTRODUCTION

1.1 Background

Collecting, compiling, analysing and publishing statistical data is a long standing method to support decision making. Statistical data is available via high-end quality data publishing platforms as well as in the form of ad-hoc created tabular data. It has to be noted that the statistical data domain was one of the first data domains that was providing open and transparent access to its data.

This value has been recognised: statistical information has been identified as "high value datasets" in the G8 Open Data Charter[1] and in its EU implementation[2]. This statement is confirmed in the Commission's Notice 2014/C 240/01[3], elaborating the results of the online consultation launched by the Commission in August 2013 on the revision to the PSI Directive[4]. According to the feedback received, statistical data was identified as one of the thematic dataset categories among those "in highest demand from re-users across the EU".

At the same time, Open Data Portals are being established throughout Europe by EU Member States. On the European level, the European Data Portal[5] has started operation in November 2015. Statistical data is of great interest for all of the data categories in such open data portals and therefore it is beneficial for references to statistical datasets to be prominently visible in such data portals.

Open data portals bring together metadata, descriptions of datasets that are hosted by data providers. The portals harvest the metadata that is provided by the providers from their content management systems in a standard exchange format. This standard metadata exchange format is known as the DCAT Application Profile for data portals in Europe (DCAT-AP)[6], developed under the aegis of the European Commission's ISA programme[7].

Through 2015, activities have already taken place towards the scoping of the work on StatDCAT-AP. Preliminary work was done by a Core Working Group with representation from Eurostat, Publications Office, DG CONNECT and representatives of ISA supported by the contractor's experts. That earlier work included definition of some terminology (data vs. metadata), an analysis of the statistical data publishing field and an analysis of standards for publishing statistical data and metadata. A

[1] Gov.uk. Cabinet Office. G8 Open Data Charter and Technical Annex. Policy paper, 18 June 2013. Action 2: Release of high value data. https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex#action-2-release-of-high-value-data
[2] European Commission. Digital Agenda for Europe. EU Implementation of the G8 Open Data Charter. 31 October 2013. http://ec.europa.eu/digital-agenda/en/news/eu-implementation-g8-open-data-charter
[3] EUR-Lex. Commission notice — Guidelines on recommended standard licences, datasets and charging for the reuse of documents. OJ C 240, 24.7.2014, p. 1–10. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52014XC0724(01)
[4] EUR-Lex. Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information. http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1449913281728&uri=CELEX:32013L0037
[5] European Commission. European Data Portal. http://www.europeandataportal.eu/
[6] European Commission. ISA – Interoperability Solutions for European Public Administrations. DCAT Application Profile for data portals in Europe. http://ec.europa.eu/isa/ready-to-use-solutions/dcat-ap_en.htm
[7] European Commission. ISA – Interoperability Solutions for European Public Administrations. http://ec.europa.eu/isa/about-isa/

conceptual mapping of SDMX to DCAT-AP was also undertaken both on the metadata level (assessing "reference" metadata created by the ESMS as the source for creation of data set descriptions) and on the data level (assessing how "structural" metadata can be derived from the data structure definition). In addition, the metadata properties used in statistical data portals such as Eurostat were evaluated.

The final report[8] of the work done in 2015 is available from the European Commission's ISA programme.

1.2 Objectives

The DCAT-AP is intended as a common layer for the exchange of metadata for a wide range of dataset types. The availability of such a common layer, creates the opportunity for a wide range of professional communities to hook onto the emerging landscape of interoperable portals by aligning with the common exchange format. In addition to the basic DCAT-AP, specific communities can extend the basic Application Profile to support description elements specific for their particular data.

The development of a DCAT-AP extension for the exchange of metadata for statistical datasets, called StatDCAT-AP, is in line with that approach, first by **determining which description elements in statistical data standards can be exposed in the DCAT-AP format**, and second by **extending the DCAT-AP with descriptive elements that can further help in the discovery and use of statistical data sets**.

The work on StatDCAT-AP is a first activity in the context of a wider roadmap of activities that aim to deliver specifications and tools that enhance interoperability between descriptions of statistical data sets within the statistical domain and between statistical data and open data portals. This roadmap, outlined in the next section, includes several activities that take place over a longer period.

The work on the specification of the StatDCAT-AP contained in this document took place over a period of eight months from November 2015 through June 2016 and covered a set of initial activities in this context. The aim of this first step was that, within the available time and resources, concrete results could be achieved that act as a demonstration and a reality check for the roadmap.

The overall objective of this first phase of work is summarised in the following charter:

[8] D02.01.2 Specification of StatDCAT-AP. A statistical extension for the DCAT application profile for data portals in Europe. Version 0.11. 2015-09-25. Available on request.

The StatDCAT-AP activity is a first step in a roadmap that aims to enhance interoperability between descriptions of statistical data sets and general data portals, facilitating referencing of statistical data with other open data.

The concrete objective of the work is to develop and reach consensus on an Application Profile of the Data Catalog Vocabulary (DCAT) to be used for the description of statistical data sets with an initial focus on discovery of those data sets in a wider context.

The StatDCAT-AP will be based on the DCAT Application Profile for Data Portals in Europe (DCAT-AP). In addition, initial guidelines on the extraction of relevant metadata from the existing implementation at Eurostat and possibly others will be elaborated in order to enable the export of metadata conforming to the application profile from existing data.

Based on the contributions of the main stakeholders, extensions to DCAT-AP can be proposed with descriptive elements particularly useful for discovery of statistical data sets beyond the possibilities offered by the generic DCAT-AP.

The work in this phase will concentrate on use cases that improve the discovery of statistical data sets published in open data portals across European institutions and EU Member States and in particular in the European Open Data Portal, as well as use cases that facilitate the integration of statistical data sets with open data from other domains.

The participants in this work had the opportunity to collaborate with colleagues from the statistical domain and with experts from the open data community, contributing and sharing their knowledge and experience with the current implementations of the statistical data standards, and were able to gain insight into possible approaches by which statistical data can be better disclosed outside of the statistical domain.

1.3 Roadmap

The wider roadmap involves several steps as listed here:

1. Connecting descriptions of statistical datasets with general open data portals through a common basic exchange format, i.e. the StatDCAT-AP;
2. Developing guidelines for the extraction of metadata from specific implementations of statistical standards towards the common exchange format;
3. Harmonising implementations of statistical standards towards a more coherent landscape of statistical resources, possibly as an extension of the basic StatDCAT profile (for the metadata level) and through the use of W3C RDF Data Cube Vocabulary (for the data level),
4. Creating a set of tools to facilitate automatic extraction and validation of metadata from data described by statistical standards into StatDCAT-AP;
5. Conducting practical pilots in various stages of the above activities to test and verify approaches and solutions.

The work reported in this document covers the first two points of the roadmap.

1.4 Structure of this document

This section 1 provides an introduction with background, objectives and roadmap.

Section 2 contains the terminology used in this document.

Section 3 presents related work, both in the statistical domain, including ongoing collaboration between Eurostat and the Publications Office of the EU, SDMX and ESMS, and in the Open Data domain, including DCAT, DCAT-AP, GeoDCAT-AP and the Data Cube vocabulary.

Section 4 outlines two use cases, one related to the improvement of the discoverability of statistical datasets on open data portals and one concerning federation of open data portals.

Section 5 describes the methodology of the work, referring to the process and methodology for the development of ISA Core Vocabularies, and outlining the analysis and decision framework, the stakeholders and the time plan for the work.

Section 6 describes the StatDCAT-AP data model in relation to the DCAT-AP model with a presentation of the elements that StatDCAT-AP adds to the DCAT-AP.

Section 7 contains a conformance statement.

Section 8 outlines the possible approaches towards export of data from existing systems into the StatDCAT-AP.

Section 9 describes the mapping of SDMX to DCAT.

Section 10 presents two options for a SDMX-based transformation mechanism, one based on SDMX structural metadata and one on SDMX Metadata Set.

Section 11 contains the acknowledgements of the people and organisations that have contributed to this work.

2 TERMINOLOGY USED IN THIS DOCUMENT

An **Application Profile** is a specification that re-uses terms from one or more base standards, adding more specificity by identifying mandatory, recommended and optional elements to be used for a particular application, as well as recommendations for controlled vocabularies to be used.

A **Dataset** is a collection of data, published or curated by a single source, and available for access or download in one or more formats.

A **Data Portal** is a Web-based system that contains a data catalogue with descriptions of datasets and provides services enabling discovery and re-use of the datasets.

Further terminology to be added.

3 RELATED WORK

3.1 Statistical data and metadata initiatives

3.1.1 Eurostat and EU Publications Office collaboration

In the context of the European Union Open Data Portal (EU ODP)[9], the Publications Office and Eurostat collaborate on the automated ingestion of Eurostat's datasets into the EU ODP. For that, there exists a mapping from the Eurostat metadata into the EU ODP metadata representation (a preliminary version of DCAT-AP)[10]. Today the Publications Office is in the transition process to align with DCAT-AP. As Eurostat is the largest contributor of datasets to EU ODP, StatDCAT-AP is a joint initiative by Eurostat and Publications Office to make more high quality metadata associated with the statistical datasets also available in a more general context of Open Data Portals.

The work is supported also by DG CONNECT, since the Pan-European data portal will be one of the key implementers of the StatDCAT-AP as the common metadata standard for harmonising the descriptions of statistical datasets originating from different countries.

The Interoperability Solutions of European Public Administrations (ISA) Programme of the European Commission is, through ISA Action 1.1, the sponsor of the activity.

3.1.2 SDMX

SDMX[11], which stands for Statistical Data and Metadata eXchange is an international initiative that aims at standardising and modernising ("industrialising") the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries.

SDMX is sponsored by seven international organisations including the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat (Statistical Office of the European Union), the International Monetary Fund (IMF), the Organisation for Economic Cooperation and Development (OECD), the United Nations Statistical Division (UNSD), and the World Bank.

These organisations are the main players at world and regional levels in the collection of official statistics in a large variety of domains (agriculture statistics, economic and financial statistics, social statistics, environment statistics etc.).

The main components of SDMX are presented in Figure 1.

[9] European Union Open Data Portal. http://open-data.europa.eu
[10] See the file ESTAT_xxx.zip in http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing
[11] Statistical Data and Metadata eXchange. https://sdmx.org/

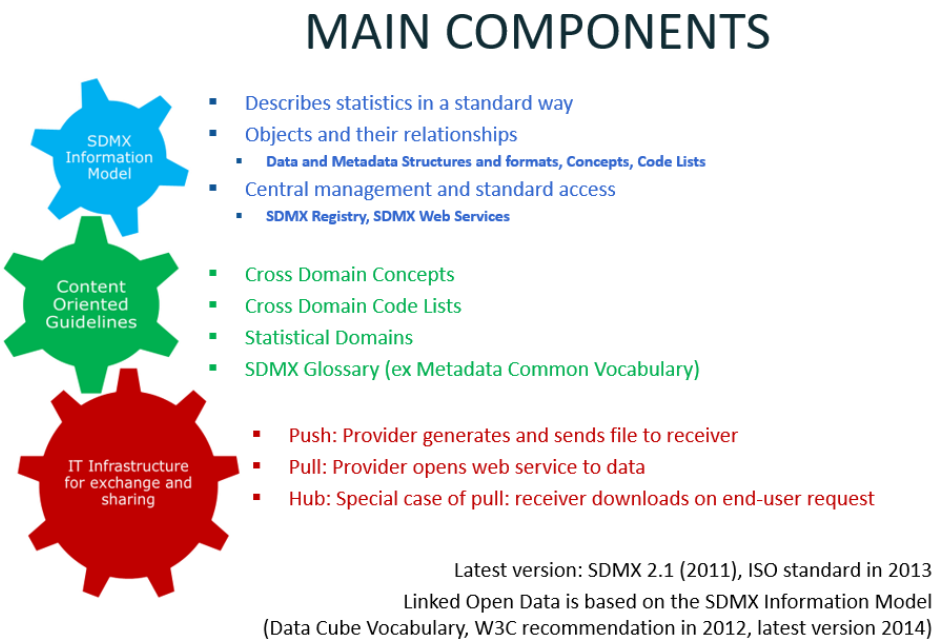


Figure 1: SDMX Main Components

A schematic view of the information model can be seen in Figure 2.

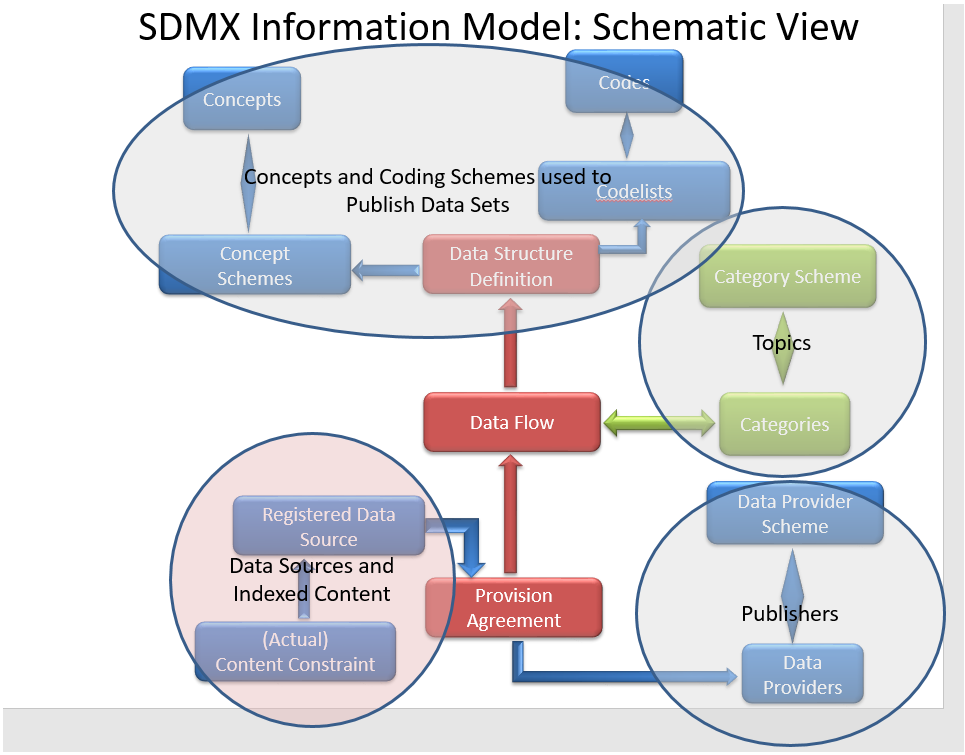


Figure 2: SDMX Information Model: Schematic View

3.1.3 ESMS

The Euro SDMX Metadata Structure (ESMS)[12] contains the description and representation of statistical metadata concepts to be used for documenting statistical data and for providing summary information useful for assessing data quality and the production process in general. The broad concepts used are based on SDMX cross-domain concepts and with the common terminology as published within the SDMX Glossary (all published in January 2009 and updated in 2016). Its structure is defined by an SDMX Metadata Structure Definition.

The ESMS is addressed to the European Statistical System and was embedded in a European Recommendation in 2009. It is implemented at Eurostat and at national level: the application of the concepts and sub concepts at European level and at national level is stated in the ESS guidelines.

The information to be entered is normally free text, but some coded elements may be introduced in the future: this is already indicated in the column "representation".

The ESMS allows the creation of different output files comprising information related to all the concepts listed or a subset of those concepts. These output files can be used for different purposes (data dissemination, quality reporting, etc.).

3.2 Open Data standards and application profiles

3.2.1 W3C DCAT

The basis for DCAT-AP is the specification of the Data Catalog Vocabulary (DCAT)[13]. DCAT was developed in the period from June 2011 through December 2013 by the Government Linked Data Working Group[14]. The specification was published as a W3C Recommendation in January 2014.

The abstract in the specification describes it as follows:

DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. This document defines the schema and provides examples for its use.

By using DCAT to describe datasets in data catalogs, publishers increase discoverability and enable applications easily to consume metadata from multiple catalogs. It further enables decentralized publishing of catalogs and facilitates federated dataset search across sites. Aggregated DCAT metadata can serve as a manifest file to facilitate digital preservation.[15]

The specification defines RDF Classes and Properties in a model that has four main entities:

- Catalogue (dcat:Catalog), defined as *a curated collection of metadata about datasets*

[12] Eurostat. Euro-SDMX Metadata Structure (ESMS). http://ec.europa.eu/eurostat/data/metadata
[13] W3C. Data Catalog Vocabulary (DCAT). W3C Recommendation 16 January 2014. http://www.w3.org/TR/vocab-dcat/
[14] W3C. Government Linked Data Working Group. https://www.w3.org/2011/gld/charter
[15] US spelling from the original.

- Catalogue Record (dcat:CatalogRecord), defined as *a record in a data catalog, describing a single dataset*
- Dataset (dcat:Dataset), defined as *a collection of data, published or curated by a single agent, and available for access or download in one or more formats*
- Distribution (dcat:Distribution), defined as representing *a specific available form of a dataset. Each dataset might be available in different forms, these forms might represent different formats of the dataset or different endpoints. Examples of distributions include a downloadable CSV file, an API or an RSS feed*

The data model of DCAT is presented in Figure 3.

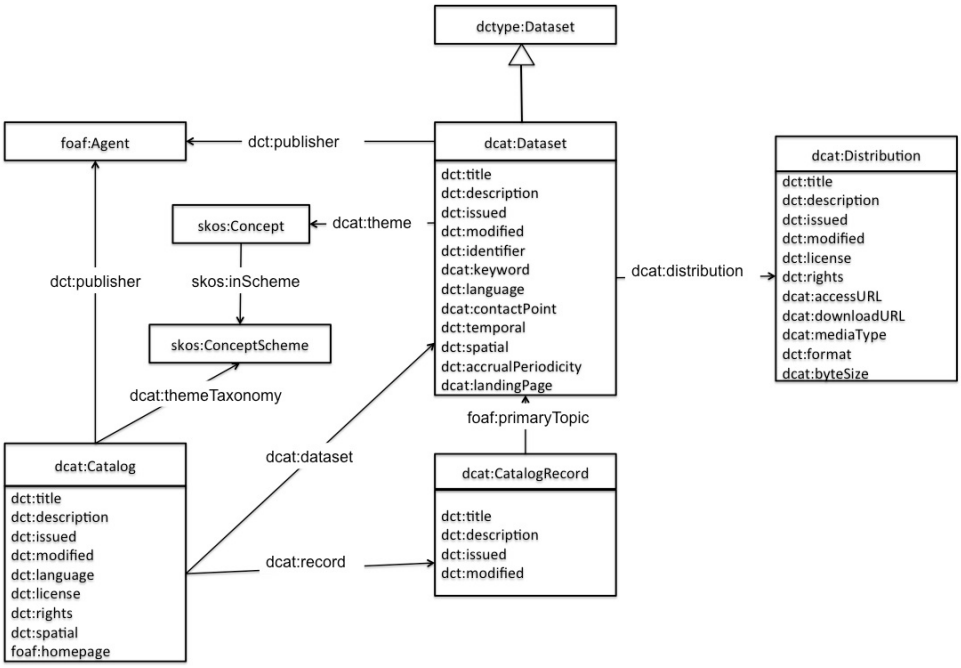


Figure 3: DCAT schematic data model

3.2.2 DCAT-AP for open data portals in Europe

The DCAT Application profile for data portals in Europe (DCAT-AP) is a specification based on W3C's Data Catalogue vocabulary (DCAT) for describing public sector datasets in Europe. Its basic use case is to enable a cross-data portal search for data sets and make public sector data better searchable across borders and sectors. This can be achieved by the exchange of descriptions of data sets among data portals.

The specification of the DCAT-AP was a joint initiative of DG CONNECT, the EU Publications Office and the ISA Programme. The specification was elaborated by a multi-disciplinary Working Group with representatives from 16 European Member States, some European Institutions and the US.

The first version (1.0)[16] of the Application Profile was published in September 2013. In 2015, a revised version (1.1)[17] was developed and published in November 2015 with changes based on requests from implementers of the first version.

The data model of DCAT-AP is presented in Figure 4.

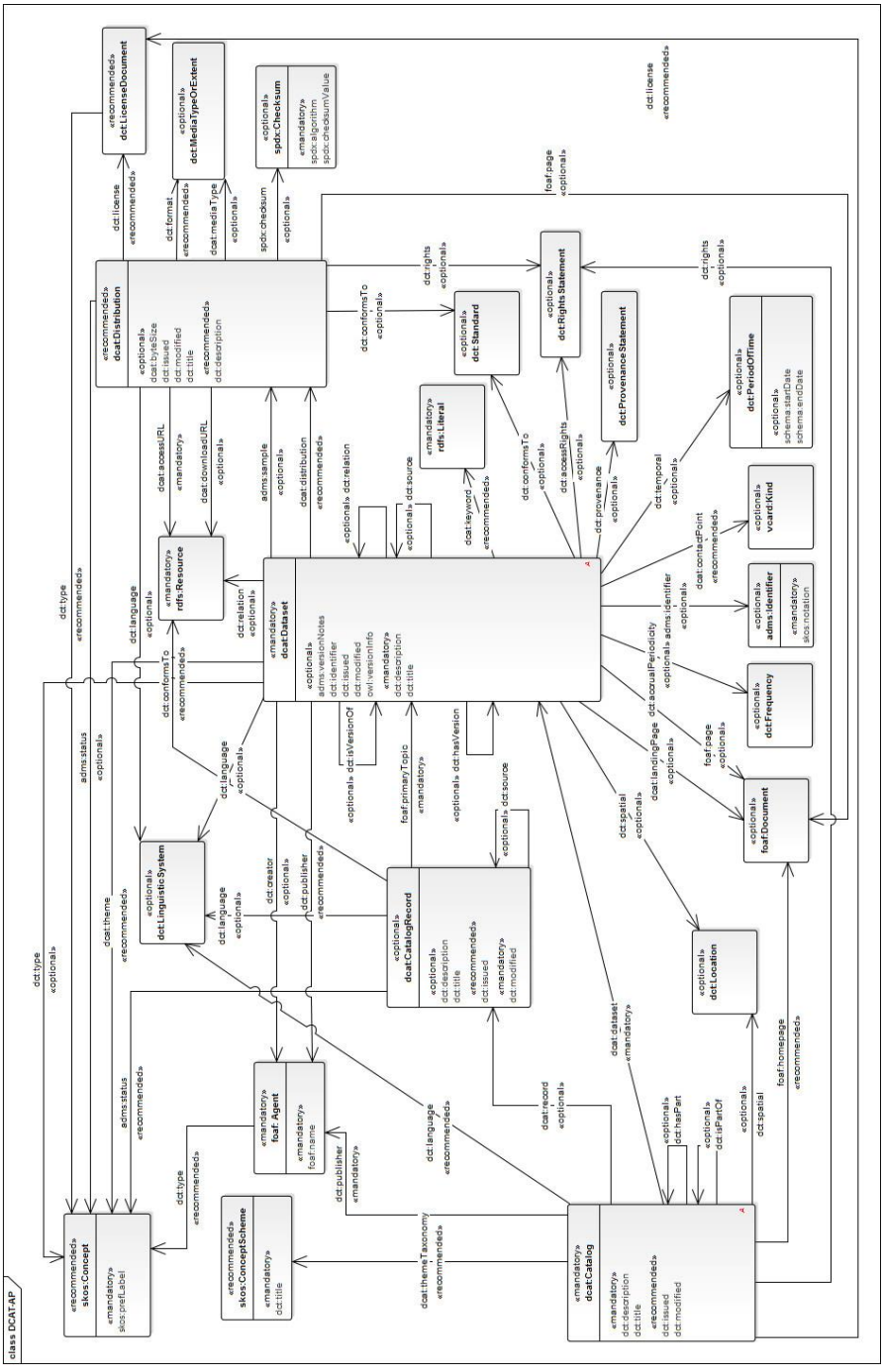


Figure 4: DCAT-AP Data Model

[16] European Commission. Joinup. DCAT application profile for data portals in Europe. Final. https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-application-profile-data-portals-europe-final

[17] European Commission. Joinup. DCAT application profile for data portals in Europe. DCAT-AP v1.1. https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-ap-v11

3.2.3 GeoDCAT-AP

GeoDCAT-AP is an extension of DCAT-AP for describing geospatial datasets, dataset series, and services. It provides an RDF syntax binding for the union of metadata elements defined in the core profile of ISO 19115:2003 and those defined in the framework of the INSPIRE Directive. Its basic use case is to make spatial datasets, data series, and services searchable on general data portals, thereby making geospatial information better searchable across borders and sectors. This can be achieved by the exchange of descriptions of data sets among data portals.

In particular, GeoDCAT-AP intends to:

- To provide an RDF syntax binding for the *union* of the elements in the INSPIRE metadata schema and the core profile of ISO 19115:2003. The guiding design principle is to make the resulting RDF syntax as simple as possible; thereby maximally using existing RDF vocabularies – such as the Dublin Core and DCAT-AP –, and as much as possible avoiding minting new terms. The defined syntax binding must enable the conversion of metadata records *from* ISO 19115 / INSPIRE *to* a harmonised RDF representation. The ability to convert metadata records from RDF to ISO 19115 / INSPIRE is not a requirement.
- To formulate recommendations to the Working Group dealing with the revision of the DCAT-AP, to maximally align DCAT-AP and GeoDCAT-AP.
- To take into account and refer to alignment of relevant controlled vocabularies (e.g., the alignments between GEMET, INSPIRE themes, EuroVoc carried out by the Publications Office of the EU[18]).

The GeoDCAT-AP specification builds upon prior work conducted by the European Commission's Joint Research Centre in 2014. This work consisted of an alignment exercise between INSPIRE metadata and DCAT-AP (version 1.0) in the framework of ISA Action 1.17 [INSPIRE-DCAT]. The results of this alignment exercise, referred to as INSPIRE+DCAT-AP, are divided in two parts:

- A Core version which defines alignments for the subset of INSPIRE metadata elements supported by DCAT-AP.
- An Extended version which defines alignments for all the INSPIRE metadata elements using DCAT-AP and other vocabularies whenever DCAT-AP is not relevant.

GeoDCAT-AP is a joint initiative of the Joint Research Centre (JRC), Unit H.6 (Digital Earth and Reference Data), the Publications Office of the European Union (PO), and the Directorates-General for Informatics (DIGIT, in the context of the ISA Programme) and Communications Networks, Content & Technology (CONNECT) of the European Commission. More than 52 people from 12 EU Member States contributed to the specification in the Working Group or during the public review period.

The first version (1.0)[19] of the GeoDCAT-AP was published in December 2015.

[18] http://publications.europa.eu/mdr/eurovoc/
[19] European Commission. Joinup. DCAT application profile for data portals in Europe. GeoDCAT-AP v1.0. https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/geodcat-ap-v10

3.2.4 The Data Cube Vocabulary

The Data Cube Vocabulary is an RDF vocabulary for representing multi-dimensional "data cubes" in RDF.

The Data Cube vocabulary provides a means to publish multi-dimensional data, such as statistics, on the web in such a way that it can be linked to related data sets and concepts using the W3C RDF (Resource Description Framework) standard. The model underpinning the Data Cube vocabulary is compatible with the cube model that underlies SDMX (Statistical Data and Metadata eXchange), an ISO standard for exchanging and sharing statistical data and metadata among organizations. The Data Cube vocabulary is a core foundation which supports extension vocabularies to enable publication of other aspects of statistical data flows or other multi-dimensional data sets.

The Data Cube vocabulary was published as a Recommendation[20] by W3C in January 2014.

An overview of its key terms and their relationships is shown in Figure 5.

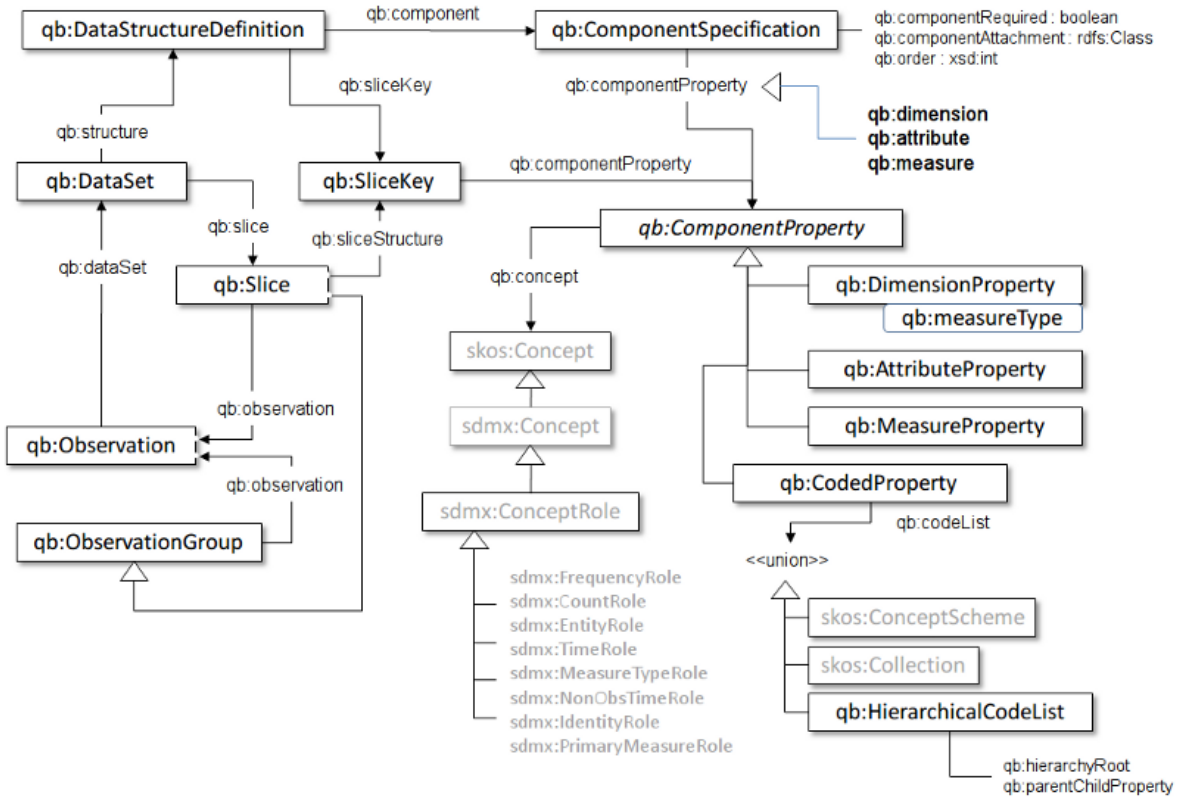


Figure 5: Data Cube vocabulary overview of key terms and relationships

[20] W3C. The RDF Data Cube Vocabulary. W3C Recommendation 16 January 2014. https://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/

4 USE CASES

4.1 Improve discoverability of statistical datasets on open data portals

Within the EU, Eurostat is the organization having as mission to provide the European Union with statistics at European level that enable comparisons between countries and regions.

In February 2015, Eurostat published more than 6500 datasets on the European Union Open Data Portal (EU ODP). That represents approximately 81% of the datasets in the European Union Open Data Portal. In practice many of the other datasets on the EU ODP are more elaborated datasets based on the datasets provided by Eurostat. On other governmental open data portals, the quantitative impact of statistical data is similarly high.

So improving the metadata quality by establishing a dedicated extended profiling of DCAT-AP, StatDCAT-AP, for statistical data has an important impact in the already published dataset records. The improvement increases public and cross-sector access to this category of high value datasets.

4.2 Federation of open data portals

At inter-institutional level, Eurostat plays an important and active role in constantly improving the exchange of statistical data. In the recent past, the world wide most prominent statistical data organizations, including Eurostat, defined and adopted the SDMX standard for the exchange of statistical data. SDMX ensures the exchange of statistical data happens without loss of information, in particular provenance information. Decision making on the sending and the receiving end of the exchange is hence based on the same information.

Open Data Portals are catalogues of dataset metadata descriptions. Within the European Union, the application profile of the W3C standard DCAT, DCAT-AP harmonizes the dataset metadata descriptions. By correlating the metadata descriptions provided by SDMX and other existing standards for statistical data, both worlds get better connected. StatDCAT-AP aims to facilitate a better integration of the existing statistical data portals with the Open Data Portals, improving the discoverability of statistical datasets.

Today Eurostat and Publications Office have established a first version of such integration. This experience and the experience gathered during work to define StatDCAT-AP can be transferred to similar setups in the EU member states.

Note that it is not the objective of StatDCAT-AP to cover actual data. For that the W3C vocabulary Data Cube[21] exists. Work on StatDCAT-AP may, however, include discussions at this level since it may improve insight.

[21] W3C. The RDF Data Cube Vocabulary. W3C Recommendation 16 January 2014 http://www.w3.org/TR/vocab-data-cube/

5 METHODOLOGY

5.1 ISA Core Vocabulary process and methodology

This work is conducted according to a process and methodology[22] that were defined for the ISA programme. The process involves the setting up of the Working Group and the publication of drafts of the specification with external review. The methodology is concerned with the elements that the specification should contain, including use cases and definition of terms and vocabularies.

The objective of the process and methodology is to involve the main stakeholders and to reach consensus in an open collaboration.

The work is conducted in a transparent manner, visible to the public through:

- A Web page
  https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/description
- An issue tracker
  https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/issue/all
- A mailing list
  http://joinup.ec.europa.eu/mailman/listinfo/stat_dcat_application_profile

5.2 Analysis and decision framework

The principle underlying the work on StatDCAT-AP are:

- align with DCAT and DCAT-AP
- focus primarily on metadata elements that contribute to discovery
- using metadata terms from existing, well-known and well-maintained vocabularies, including ISA Core Vocabularies
- encourage the use of common controlled vocabularies, preferably ones maintained in MDR by Publications Office
- find an appropriate balance between simplicity and complexity from the perspective of the widest, non-specialist audience

5.3 Stakeholders

The main stakeholders of this work are:

- Eurostat
- The Publications Office of the EU
- National and regional statistical offices

In addition, the organisations responsible for operating general data portals that are interested in collecting and integrating statistical datasets in their services.

[22] European Commission. Joinup. Process and methodology for developing semantic agreements. https://joinup.ec.europa.eu/community/core_vocabularies/document/process-and-methodology-developing-semantic-agreements

5.4 Time plan

Target dates	Event, outcome
December 2015	invitations to stakeholders, set up collaboration infrastructure
January 2016	collect requirements and suggestions
5 February 2016	familiarisation Webinar
February 2016	first draft based on initial analysis and issues raised
11 March 2016	first virtual WG meeting to discuss first draft
15 April 2016	second virtual WG meeting to discuss draft mapping and implementation options
6 May 2016	second draft available for review, incorporating comments and further development
13 May 2016	third meeting (face-to-face plus Adobe Connect) in Rome to discuss mapping issues in practice
End of May 2016	third draft, including full mapping proposal and usage of controlled vocabularies
3 June 2016	fourth virtual WG meeting to agree schedule for public review
July and August 2016	public review period
Mid-September 2016	fifth virtual WG meeting to discuss and resolve public comments received
End of September 2016	approval of StatDCAT-AP version 1 for publication

6 THE STATDCAT-AP DATA MODEL

6.1 Informal description

The StatDCAT Application Profile is an extension of the DCAT Application Profile for Data Portals in Europe, version 1.1 (DCAT-AP)[23].

Its purpose is to provide a specification that is fully conformant with DCAT-AP version 1.1 as it meets all obligations of the DCAT-AP Conformance Statement. As a result, data portals that comply with DCAT-AP will be able to understand the core of StatDCAT-AP. In addition, StatDCAT-AP defines a small number of additions to the DCAT-AP model that are particular relevant for statistical datasets. Given that there are many statistical datasets that are of interest to the general data portals and their users, it is likely that recognising and exposing the additions to DCAT-AP proposed by StatDCAT-AP will be beneficial for the general data portals to be able to provide enhanced services for collections of statistical data.

The StatDCAT-AP data model includes the four main entities that are also present in DCAT-AP (see also Figure 4 for a diagram of the DCAT-AP data model):

1. The **Catalogue**: this represents a collection of Datasets. It is defined in the DCAT Recommendation[24] as "*a curated collection of metadata about datasets*". The description of the Catalogue includes links to the metadata for each of the Datasets that are in the Catalogue.
2. The **Catalogue Record**: DCAT defines this as "*a record in a data catalog, describing a single dataset*". The Catalogue Record enables statements about the description of a Dataset rather than about the Dataset itself. Catalogue Records may not be used by all implementations. It is optional in DCAT-AP and mostly used by aggregators to keep track of harvesting history.
3. The **Dataset**: this represents the published information. It is defined as "*a collection of data, published or curated by a single agent, and available for access or download in one or more formats*". The description of a Dataset includes links to each of its Distributions, if they are available. A Dataset is not required to have a Distribution; examples are Datasets that are described before the associated data is collected, Datasets for which the data has been removed, and Datasets that are only accessible through a landing page.
4. The **Distribution**: this, according to DCAT, "*represents a specific available form of a dataset. Each dataset might be available in different forms, these forms might represent different formats of the dataset or different endpoints. Examples of distributions include a downloadable CSV file, an API or an RSS feed*". The description of a Distribution contains information about the location of the data files or access point and about the file format and licence for use or reuse. In the case of statistical datasets, Distributions may be available in specific formats like SDMX-ML or using the Data Cube vocabulary.

[23] European Commission. Joinup. DCAT application profile for data portals in Europe. DCAT-AP v1.1. https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-ap-v11
[24] W3C. Data Catalogue Vocabulary. W3C Recommendation 16 January 2014. https://www.w3.org/TR/vocab-dcat/

6.2 Extensions for description of statistical datasets

The following elements are defined in this StatDCAT-AP specification in addition to the elements already present in DCAT-AP:

- Number of observations
- Number of data series
- Link to visualisation
- Dimensions as property
- Dimensions as keywords
- Quality aspects
- Statistical unit
- Statistical population

These additions are further described in section 6.7.

6.3 Overview of the model

In the following sections, classes and properties are grouped under headings 'mandatory', 'recommended' and 'optional'. These terms have the following meaning.

- **Mandatory class**: a receiver of data MUST be able to process information about instances of the class; a sender of data MUST provide information about instances of the class.
- **Recommended class**: a sender of data SHOULD provide information about instances of the class; a sender of data MUST provide information about instances of the class, if such information is available; a receiver of data MUST be able to process information about instances of the class.
- **Optional class**: a receiver MUST be able to process information about instances of the class; a sender MAY provide the information but is not obliged to do so.
- **Mandatory property**: a receiver MUST be able to process the information for that property; a sender MUST provide the information for that property.
- **Recommended property**: a receiver MUST be able to process the information for that property; a sender SHOULD provide the information for that property if it is available.
- **Optional property**: a receiver MUST be able to process the information for that property; a sender MAY provide the information for that property but is not obliged to do so.

The meaning of the terms MUST, MUST NOT, SHOULD and MAY in this section and in the following sections are as defined in RFC 2119[25].

In the given context, the term "processing" means that receivers must accept incoming data and transparently provide these data to applications and services. It does neither imply nor prescribe what applications and services finally do with the data (parse, convert, store, make searchable, display to users, etc.).

[25] IETF. RFC 2119. Key words for use in RFCs to Indicate Requirement Levels. http://www.ietf.org/rfc/rfc2119.txt

6.4 Namespaces

The Application Profile reuses terms from various existing specifications. Classes and properties specified in the next sections have been taken from the following namespaces:

- adms: http://www.w3.org/ns/adms#
- dcat: http://www.w3.org/ns/dcat#
- dct: http://purl.org/dc/terms/
- foaf: http://xmlns.com/foaf/0.1/
- owl: http://www.w3.org/2002/07/owl#
- rdfs: http://www.w3.org/2000/01/rdf-schema#
- schema: http://schema.org/
- skos: http://www.w3.org/2004/02/skos/core#
- spdx: http://spdx.org/rdf/terms#
- xsd: http://www.w3.org/2001/XMLSchema#
- vcard: http://www.w3.org/2006/vcard/ns#

6.5 UML Class diagram

Diagram; copy of diagram from DCAT-AP with highlighted extensions.

6.6 Description of classes

6.6.1 Mandatory Classes

Class name	Usage note for the Application Profile	URI	Reference
Agent	An entity that is associated with Catalogues and/or Datasets. If the Agent is an organisation, the use of the Organization Ontology[26] is recommended.	foaf:Agent	http://xmlns.com/foaf/spec/#term_Agent , http://www.w3.org/TR/vocab-org/
Catalogue	A catalogue or repository that hosts the Datasets being described.	dcat:Catalog	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog
Dataset	A conceptual entity that represents the information published.	dcat:Dataset	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-dataset
Literal	A literal value such as a string or integer; Literals may be typed, e.g. as a date according to xsd:date. Literals that contain human-readable text have an optional language tag as defined by BCP 47[27].	rdfs:Literal	http://www.w3.org/TR/rdf-concepts/#section-Literals
Resource	Anything described by RDF.	rdfs:Resource	http://www.w3.org/TR/rdf-schema/#ch_resource

6.6.2 Recommended Classes

Class name	Usage note for the Application Profile	URI	Reference
Category	A subject of a Dataset.	skos:Concept	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-category-and-category-scheme

[26] W3C. The Organization Ontology. W3C Candidate Recommendation, 25 June 2013. http://www.w3.org/TR/2013/CR-vocab-org-20130625/
[27] IETF. BCP 47. Tags for Identifying Languages. http://www.rfc-editor.org/rfc/bcp/bcp47.txt

Class name	Usage note for the Application Profile	URI	Reference
Category scheme	A concept collection (e.g. controlled vocabulary) in which the Category is defined.	skos:ConceptScheme	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-category-and-category-scheme
Distribution	A physical embodiment of the Dataset in a particular format.	dcat:Distribution	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-distribution
Licence document	A legal document giving official permission to do something with a resource.	dct:LicenseDocument	http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms#LicenseDocument

The class 'Distribution' is classified as 'Recommended' to allow for cases that a particular Dataset does not have a downloadable Distribution, and in such cases the sender of data would not be able to provide this information. However, it can be expected that in the majority of cases Datasets do have downloadable Distributions, and in such cases the provision of information on the Distribution is mandatory.

6.6.3 Optional Classes

Class name	Usage note for the Application Profile	URI	Reference
Catalogue Record	A description of a Dataset's entry in the Catalogue.	dcat:CatalogRecord	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog-record
Checksum	A value that allows the contents of a file to be authenticated. This class allows the results of a variety of checksum and cryptographic message digest algorithms to be represented.	spdx:Checksum	http://spdx.org/rdf/terms#Checksum
Document	A textual resource intended for human consumption that contains information, e.g. a web page about a Dataset.	foaf:Document	http://xmlns.com/foaf/spec/#term_Document
Frequency	A rate at which something recurs, e.g. the publication of a Dataset.	dct:Frequency	http://dublincore.org/documents/dcmi-terms/#terms-Frequency
Identifier	An identifier in a particular context, consisting of the string that is the identifier; an optional identifier for the identifier scheme; an optional identifier for the version of the identifier scheme; an optional identifier for the agency that manages the identifier scheme	adms:Identifier	http://www.w3.org/TR/vocab-adms/#identifier
Kind	A description following the vCard specification, e.g. to provide telephone number and e-mail address for a contact point. Note that the class Kind is the parent class for the four explicit types of vCards (Individual, Organization, Location, Group).	vcard:Kind	http://www.w3.org/TR/2014/NOTE-vcard-rdf-20140522/#d4e181
Linguistic system	A system of signs, symbols, sounds, gestures, or rules used in communication, e.g. a language	dct:LinguisticSystem	http://dublincore.org/documents/dcmi-terms/#terms-LinguisticSystem
Location	A spatial region or named place. It can be represented using a controlled vocabulary or with geographic coordinates. In the latter case, the use of the Core Location Vocabulary[28] is recommended, following the approach described in the GeoDCAT-AP specification.	dct:Location	http://dublincore.org/documents/dcmi-terms/#terms-Location
Media type or extent	A media type or extent, e.g. the format of a computer file	dct:MediaTypeOrExtent	http://dublincore.org/documents/dcmi-terms/#terms-MediaTypeOrExtent
Period of time	An interval of time that is named or defined by its start and end dates.	dct:PeriodOfTime	http://dublincore.org/documents/dcmi-terms/#terms-PeriodOfTime

[28] European Commission. Joinup. Core Location Vocabulary. https://joinup.ec.europa.eu/asset/core_location/description

Publisher type	A type of organisation that acts as a publisher	skos:Concept	http://www.w3.org/TR/vocab-adms/#dcterms-type
Rights statement	A statement about the intellectual property rights (IPR) held in or over a resource, a legal document giving official permission to do something with a resource, or a statement about access rights.	dct:RightsStatement	http://dublincore.org/documents/dcmi-terms/#terms-RightsStatement
Standard	A standard or other specification to which a Dataset or Distribution conforms	dct:Standard	http://dublincore.org/documents/dcmi-terms/#terms-Standard
Status	An indication of the maturity of a Distribution or the type of change of a Catalogue Record.	skos:Concept	http://www.w3.org/TR/vocab-adms/#status

6.7 Description of properties per class

The basis for StatDCAT-AP is the specification of DCAT-AP. In the next section, only the proposals for additional properties are included. In further versions, more details need to be added based on the outcomes of the decision on the proposed extensions.

StatDCAT-AP respects the conformance requirements defined for DCAT-AP version 1.1, which means that it will have, at least, the same mandatory classes and mandatory properties as DCAT-AP 1.1. StatDCAT-AP may extend DCAT-AP by specifying additional properties, as long as they are reused from existing RDF vocabularies.

The next sections present the suggested extensions which are under discussion at https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/issue/all.

6.7.1 Proposed extension: Number of observations

During discussion with stakeholders, the following additional property was proposed:

'Number of observations' as a property for Dataset

The number of observations provides information on the total number of values that are contained in the Dataset.

This property is intended to provide an indication of the size of a Dataset. DCAT-AP has an option to indicate the size in bytes of a data file through the property byteSize (https://www.w3.org/TR/vocab-dcat/#Property:distribution_size) for Distribution but that only gives the physical size of the dataset which is not the only aspect of interest for statistical Datasets.

The expected value for this property is a string in an agreed format, e.g. "20 observations".

See https://joinup.ec.europa.eu/node/151139.

6.7.2 Proposed extension: Number of data series

During discussion with stakeholders, the following additional property was proposed:

'Number of data series' as a property for Dataset

The number of data series provides information on how values in the Dataset are grouped; for example, a Dataset could contain data for three regions with three

values for each region. In this example, the number of series is three while the number of observations is nine.

This property is intended to provide an indication of the size of a Dataset. DCAT-AP has an option to indicate the size in bytes of a data file through the property byteSize (https://www.w3.org/TR/vocab-dcat/#Property:distribution_size) for Distribution but that only gives the physical size of the dataset which is not the only aspect of interest for statistical Datasets.

The expected value for this property is a string in an agreed format, e.g. "5 series".

https://joinup.ec.europa.eu/node/151140

6.7.3 Proposed extension: Link to visualisation

During discussion with stakeholders, the following additional property was proposed:

'Visualisation' as a property for Dataset

This property is intended to provide a link to a page where the data can be seen in a graphical representation. The expected value for this property is a URL that opens the visualisation for this Dataset.

https://joinup.ec.europa.eu/node/151141

6.7.4 Proposed extension: Dimensions as property

During discussion with stakeholders, the following additional property was proposed:

'Dimensions' as a property for Dataset.

This property is intended to provide insight on the level of the dataset description of the kinds of observations that are contained in the data, e.g. whether the data contains observations related to time periods, regions, sex etc. The expected value for this property is a URI of a SKOS concept scheme or code list (see https://www.w3.org/TR/vocab-data-cube/#schemes).

Participants in this activity are asked to respond to the following questions:

https://joinup.ec.europa.eu/node/151142

6.7.5 Proposed extension: Dimensions as keywords

During discussion with stakeholders, the following recommendation was proposed for the property keyword of Dataset.

It is proposed that the property 'keyword' is used to expose the 'dimensions' of the observations contained in the Dataset.

The expected value for this property is a text string with the name of a SKOS concept scheme or code list (see https://www.w3.org/TR/vocab-data-cube/#schemes).

https://joinup.ec.europa.eu/node/151143

6.7.6 Proposed extension: Quality aspects

During discussion with stakeholders it was suggested that it would be useful if quality aspects of datasets could be expressed.

https://joinup.ec.europa.eu/node/151144

6.7.7 Statistical unit

During discussion with stakeholders, the following additional property was proposed:

'Statistical unit' as a property for Dataset.

This property is intended to provide information defined in the EURO-SDMX Metadata Structure (ESMS) as the "entity for which information is sought and for which statistics are ultimately compiled". According to the ESMS guideline the property should "list the basic units of statistical observation for which data are provided. These observation units (e.g. the enterprise, the local unit, private households,...) can be different from the reporting units used in the underlying statistical surveys". The expected value for this property is text.

6.7.8 Statistical population

During discussion with stakeholders, the following additional property was proposed:

'Statistical population' as a property for Dataset.

This property is intended to provide information defined in the EURO-SDMX Metadata Structure (ESMS) as the "total membership or population or "universe" of a defined class of people, objects or events". According to the ESMS guideline the property should "describe the target statistical population (one or more) which the data set refers to, i.e. the population about which information is to be sought". The expected value for this property is text.

6.8 Controlled vocabularies

List of controlled vocabularies to be used in StatDCAT-AP descriptions, as much as possible applying vocabularies used in the general DCAT-AP, augmented by appropriate vocabularies of common usage in the statistical domain.

7 CONFORMANCE STATEMENT

7.1 Provider requirements

In order to conform to this Application Profile, an application that provides metadata MUST:

- Provide a description of the Catalogue, including at least the mandatory properties specified for this class.
- Provide information for the mandatory properties specified for the Catalogue Record class, if descriptions of Catalogue Records are provided – please note that the provision of descriptions of Catalogue Records is optional.
- Provide descriptions of Datasets in the Catalogue, including at least the mandatory properties specified for this class.
- Provide descriptions of Distributions, if any, of Datasets in the Catalogue, including at least the mandatory properties specified for this class.
- Provide descriptions of all organisations involved in the descriptions of Catalogue and Datasets, including at least the mandatory properties specified for the Agent class.
- Provide descriptions of all category schemes that contain the categories that are asserted in any of the descriptions of Datasets in the Catalogue, including at least the mandatory properties specified for the Category Scheme class.
- Provide descriptions of all categories involved in the descriptions of Datasets in the Catalogue, including at least the mandatory properties specified for the Category class.

For the properties listed in the table in section 6.8 Controlled vocabularies, the associated controlled vocabularies MUST be used. Additional controlled vocabularies MAY be used.

In addition to the mandatory properties, any of the recommended and optional properties defined for any of the classes MAY be provided.

Recommended and optional classes may have mandatory properties, but those only apply if and when an instance of such a class is present in a description.

7.2 Receiver requirements

In order to conform to this Application Profile, an application that receives metadata MUST be able to:

- Process information for all classes specified.
- Process information for all properties specified.
- Process information for all controlled vocabularies specified.

In this context, "processing" means that receivers must accept incoming data and transparently provide these data to applications and services. It does neither imply nor prescribe what applications and services finally do with the data (parse, convert, store, make searchable, display to users, etc.).

8 MAPPING AND EXTRACTION APPROACHES

As there will be no systems that implement StatDCAT-AP as a native format. As the StatDCAT-AP format is intended as a common target format for export of data that may exist in a variety of standard and local formats, in many modelled in SDMX, the provision of information according to this StatDCAT-AP specification will involve some form of extraction or mapping process.

The approach to this extraction of mapping will be dependent on the local data structures and technical environment and this document does not prescribe the way that local implementers may want to build the necessary extraction and mapping mechanisms. This is entirely the responsibility of the local implementers.

While it is likely that there will be cases where the export to StatDCAT-AP is done directly from the local structures, it might also be helpful for implementers that manage local systems that are based on SDMX, to map their data to a SDMX-based intermediary format.

Such a format may enable common approaches among SDMX implementations and may lower the threshold for the export of data conformant to StatDCAT-AP from SDMX-based systems.

So, while for implementers that opt for directly exporting StatDCAT-AP from local formats, the specification of StatDCAT-AP in section 6 is all they need to develop their extraction and mapping modules, SDMX implementers may consider basing their work on the approaches presented in the following sections 9 Mapping SDMX to DCAT and 10 SDMX-based Transformation Mechanism.

9 MAPPING SDMX TO DCAT

9.1 Scope

The scope of this section is to describe the mapping of DACT-AP to the SDMX Information Model. This is achieved by means of schematic diagrams of the SDMX Information Model and also by a worked example where the SDMX-ML content is mapped to the classes and properties of DCAT-AP.

The intent of this mapping is twofold:

1. It enables those organisations that are using SDMX to know which metadata structures to use in order to create DCAT-AP directly from their SDMX metadata repository (such as an SDMX Registry).
2. It enables organisations, other than those mentioned in (1) above, that wish to use SDMX-ML structural metadata as the format for the Transformation Mechanism (described in Section 10 of this specification), to know which SDMX-ML element or attribute maps to which DCAT-AP class or property.

9.2 Diagrams

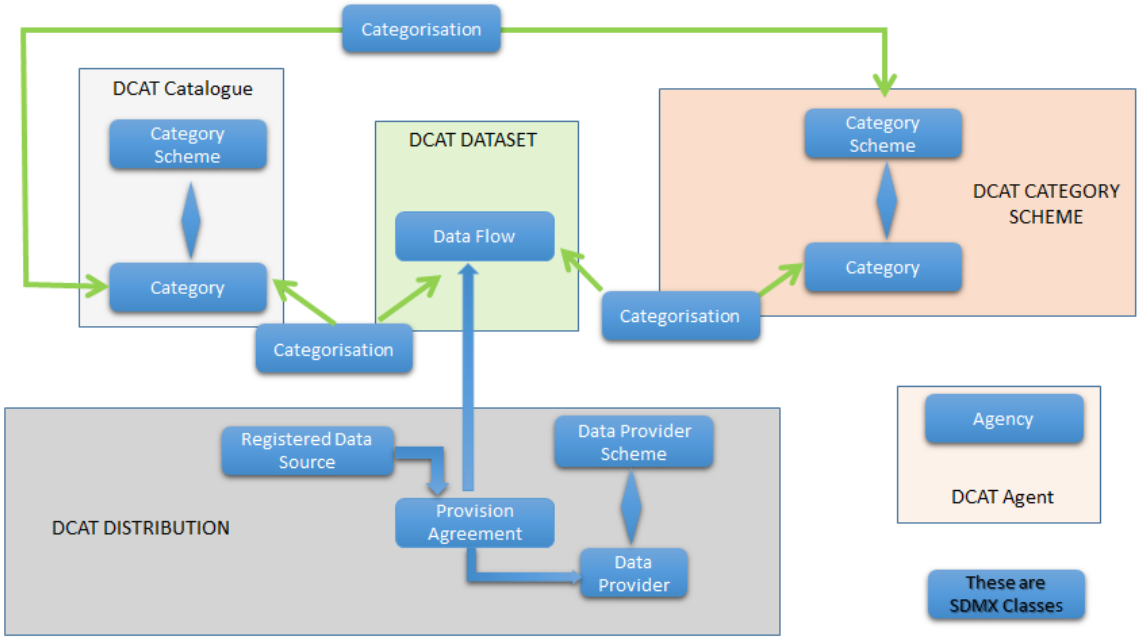


Figure 6: Schematic map of SDMX Classes to DCAT-AP

This is a schematic diagram of those high level classes in the SDMX Information Model that provide the metadata required by StatDCAT-AP.

A narrative explanation is:

1) The DCAT Catalogue is mapped to an SDMX Category Scheme. The Category can link to any other structural metadata object in SDMX using a Categorisation. The Categorisation provides the link i.e. the Categorisation references both the object and the Category to which it is linked. Two

Categories are present in the Category Scheme representing the DCAT-Catalogue, one for linking the Dataflows, and one for linking the Category Scheme containing the topic themes. There will be multiple Categorisations, each one linking the object (e.g. Dataflow) to the relevant Category. Therefore, for instance, there will be one Categorisation for each Dataflow, each Categorisation referencing the same Category. In this way all of the Dataflows that are contained in the catalogue are linked to the same Category.

2) The DCAT Dataset maps to the SDMX Dataflow.

3) The DCAT Category Scheme maps to the SDMX Category Scheme. Note that this will be a different physical Category Scheme to the one that contains the DCAT Catalogue. The Categories in this Category Scheme are the topics or themes that categorise the type of data. Each Category links to the Dataflows that are relevant to the topic by means of a Categorisation. A Dataflow may be linked to many such topics (Categories) and a topic (Category) can be linked to many Dataflows.

4) The DCAT Distribution maps to the SDMX Provision Agreement which links a Data Provider with a Dataflow. The Data Provider and the Dataflow have a many-to-many association, each one-to-one association is represented as a Provision Agreement. The actual data source for one Data Provider and its linked Dataflow is the Registered Data Source linked to the Provision Agreement. The URL of the Registered Data Source is a link to a data source, which can be a URL that resolves to an actual set of data or it may be a URL to a web service that can be queried for the data. SDMX makes a distinction between the two.

5) The DCAT Agent maps to the SDMX Agency which is the "Maintenance Agency" for the metadata such as the Dataflow. Note that in SDMX the Maintenance Agency is maintained in a different scheme to the Data Provider. So, the Data Provider is a different construct from the Agency. In SDMX the Data Provider (of the actual data) can be different from the Maintenance Agency of the metadata describing the data (the SDMX Dataflow): they may both have the same Id but are different entities.

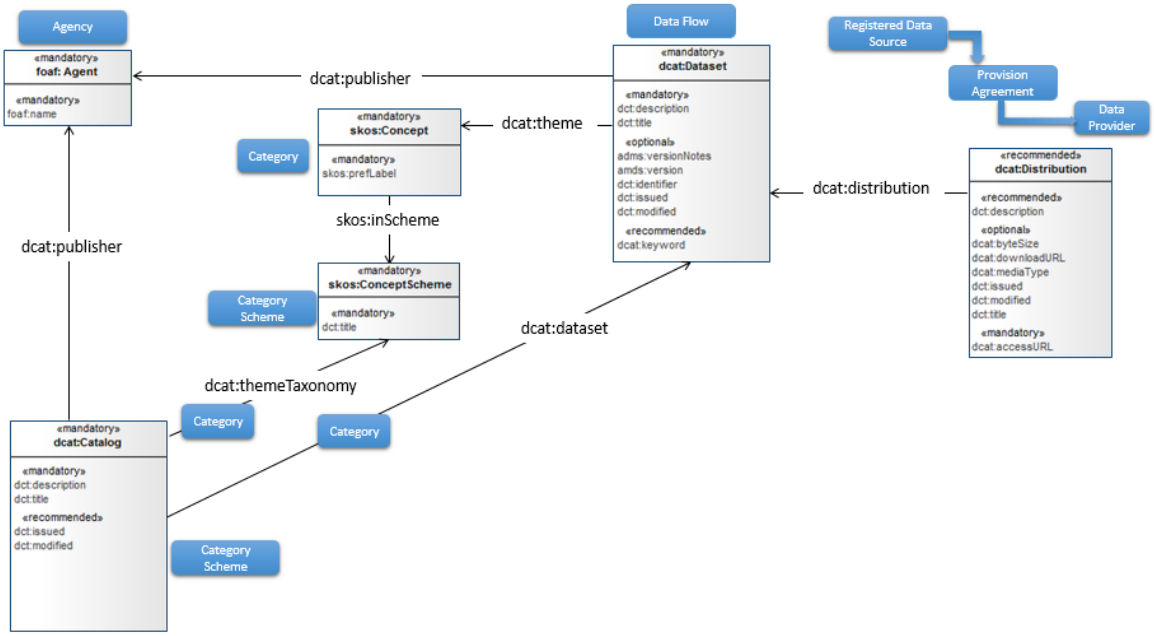


Figure 7: DCAT-AP Model mapped to SDMX Model Classes

This shows the same mapping but from the perspective of the DCAT-AP model.

9.3 Example

9.3.1 Introduction

This example shows how the SDMX structural metadata are mapped to the DCAT-AP classes and properties. The mapping shows the XML instances of the structural metadata authored in an SDMX Registry and exported as SDMX-ML.

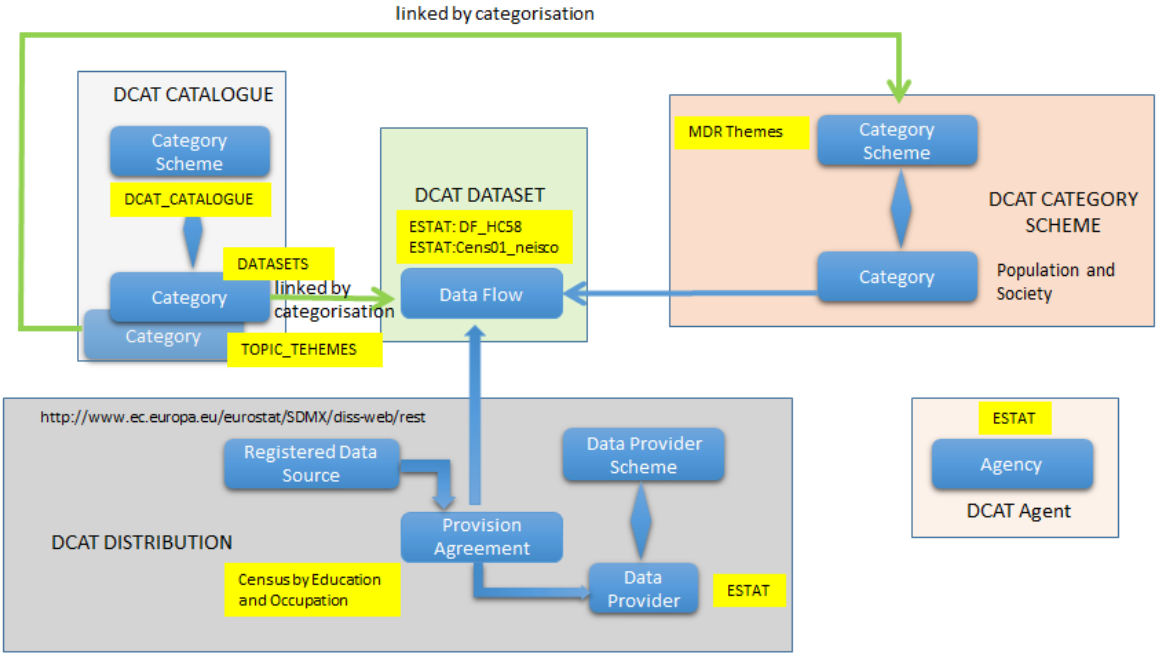


Figure 8: Metadata Used in the Example Mapping

This shows the schematic diagram of the high level SDMX classes and the content of these for the instance of these classes used in the examples that follow.

A narrative explanation is:

1) The SDMX Category Scheme containing the DCAT Catalogue has two Categories. One Category (*TOPIC_SCHEMES*) links to the DCAT Category Scheme of *MDR Themes*, the other Category (*DATASETS)* links to all of the DCAT Datasets (in SDMX this is called a Dataflow) contained in the Catalogue.
2) In the example two Dataflows are present: *DF_HC58* (Census Hub Hypercube 58) and *Cens01_neisco* (Census data broken down by education and occupation). Both Dataflows are maintained by Eurostat (Agent=*ESTAT*). The Dataflow *DF_HC58* is included only to show how the SDMX Category can link to multiple Dataflows. The Dataflow *Cens01_neisco* is the one used for the detailed mapping of SDMX to the StatDCAT-AP classes – Dataset, Distribution, Category Scheme, Agent
3) The SDMX Category Scheme containing the list of has the name *MDR Themes* in the examples.
4) The Provision Agreement containing the DCAT Distribution in the example is named *Census by Education and Occupation* and links the Data Provider (*ESTAT)* to the Dataflow *Cens01_neisco*. The URL of the Registered Data Source is a link to a web service that can be queried for the data.
5) The SDMX Agency containing the DCAT Agent is *ESTAT.* The Data Provider is a different construct from the *ESTAT* Agency, but in this example it is given the same Id (*ESTAT*).
6) The URL in the Registered Data Source (*http://www.ec.europa.eu/eurostat/SDMX/diss-web/rest*) is the dcat:accessURL in the DCAT Distribution.

9.3.2 SDMX Annotations

SDMX does not support some of the mandatory or recommended properties of DCAT-AP. However, SDMX has an extensibility mechanism called "Annotations". Annotations can be added to any SDMX object that can be identified.

The structure of an Annotation is shown below:

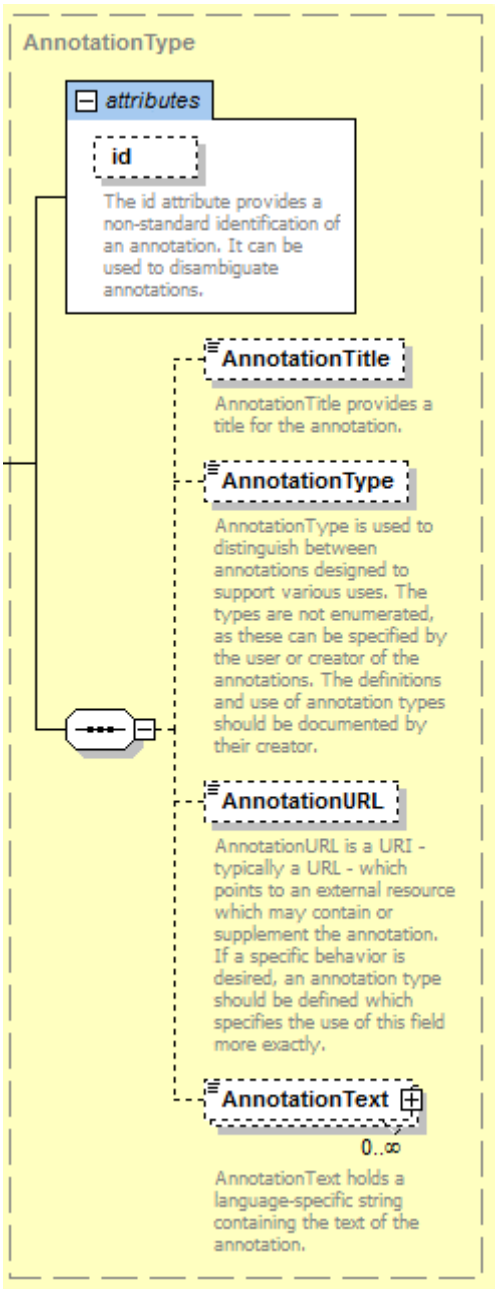


Figure 9: SDMX XML schema specification for Annotation

In the examples that follow the following elements are used:

AnnotationTitle contains the DCAT-AP property value

AnnotationType contains the value StatDCAT-AP indicating that this is a StatDCAT-AP property

AnnotationURL is a URI

AnnotationText is a text value (this can occur many times to support be multilingual variants)

9.3.3 Data Catalogue

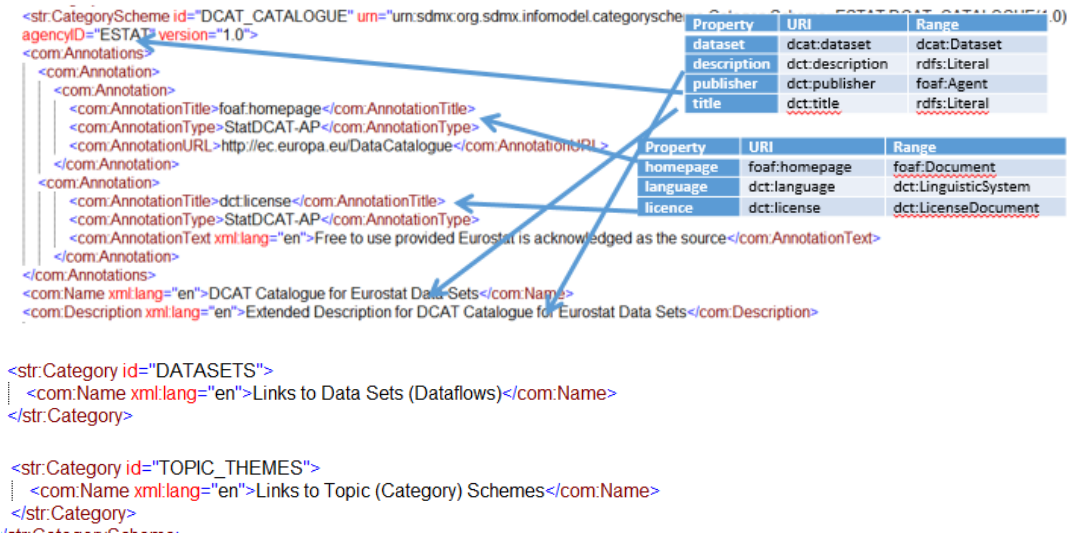
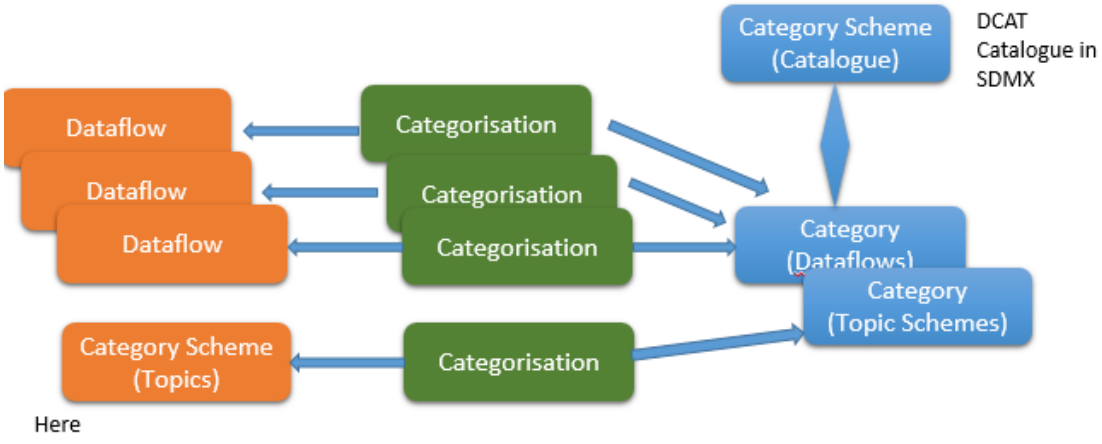


Figure 10: SDMX-DCAT mapping example for the DCAT Catalogue

9.3.4 Linking to Categories using Categorisations

Schematic



Here

- One Category links to all of the Dataflows (Datasets) of the DCAT catalogue
- One category links to the Category Scheme of (DCAT) topics (can link to many of these if required)

Figure 11: Schematic showing linking of SDMX Categories to other SDMX objects

Example

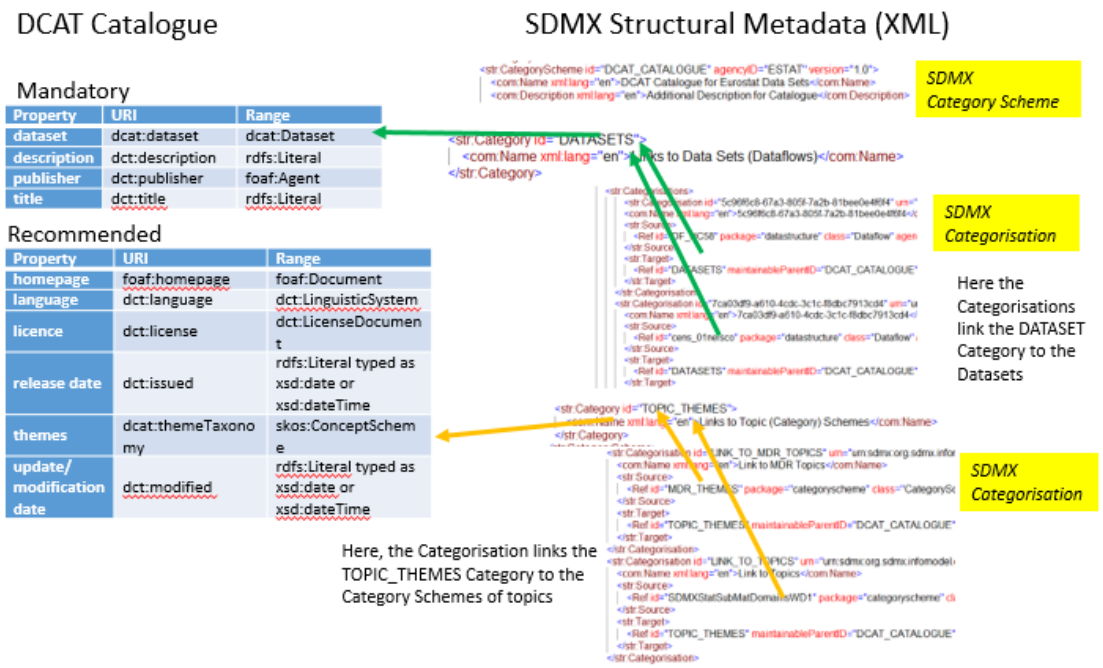


Figure 12: Linking Catalogue to DCAT Datasets and Category (Topic) Scheme

9.3.5 Dataset

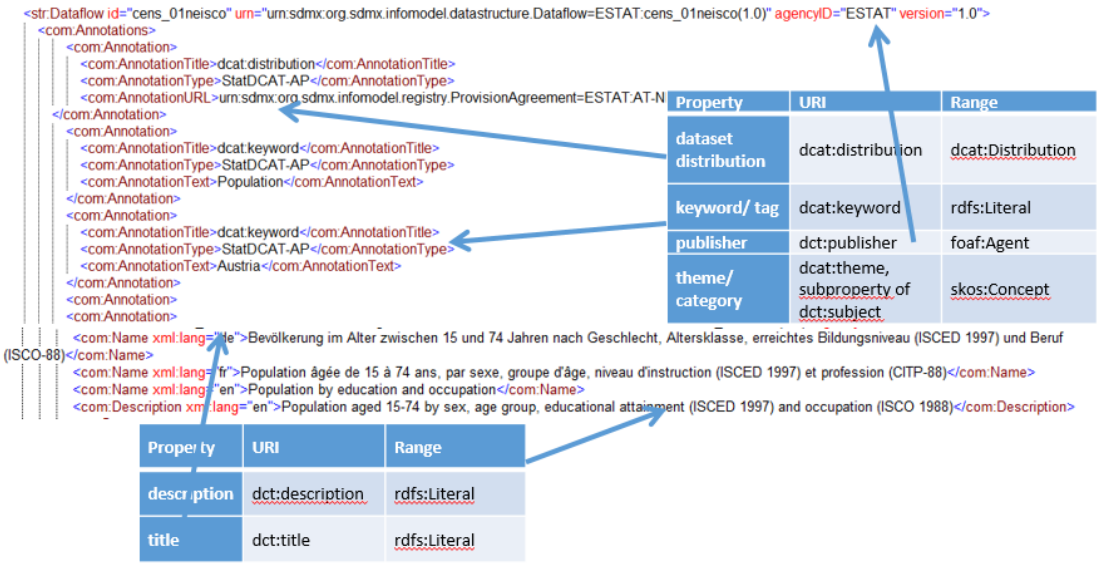


Figure 13: SDMX-DCAT mapping example for the DCAT Dataset

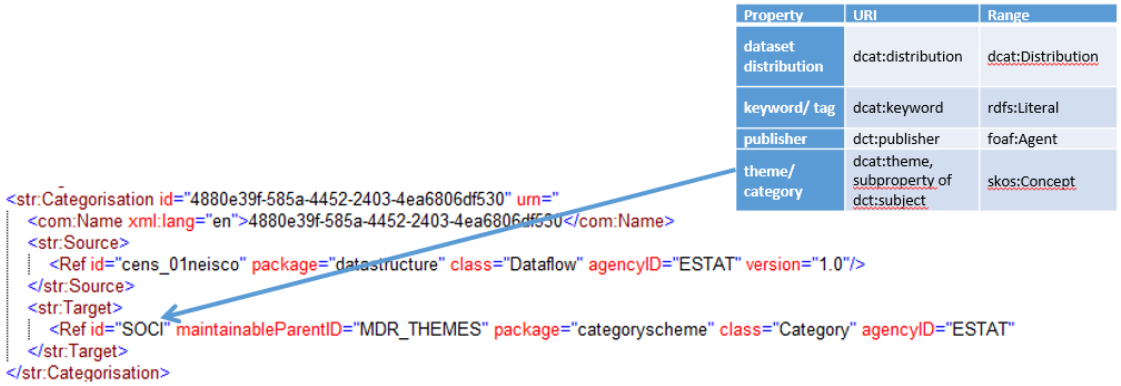


Figure 14: Linking a Dataflow to the SDMX Category (Topic)

9.3.6 Distribution

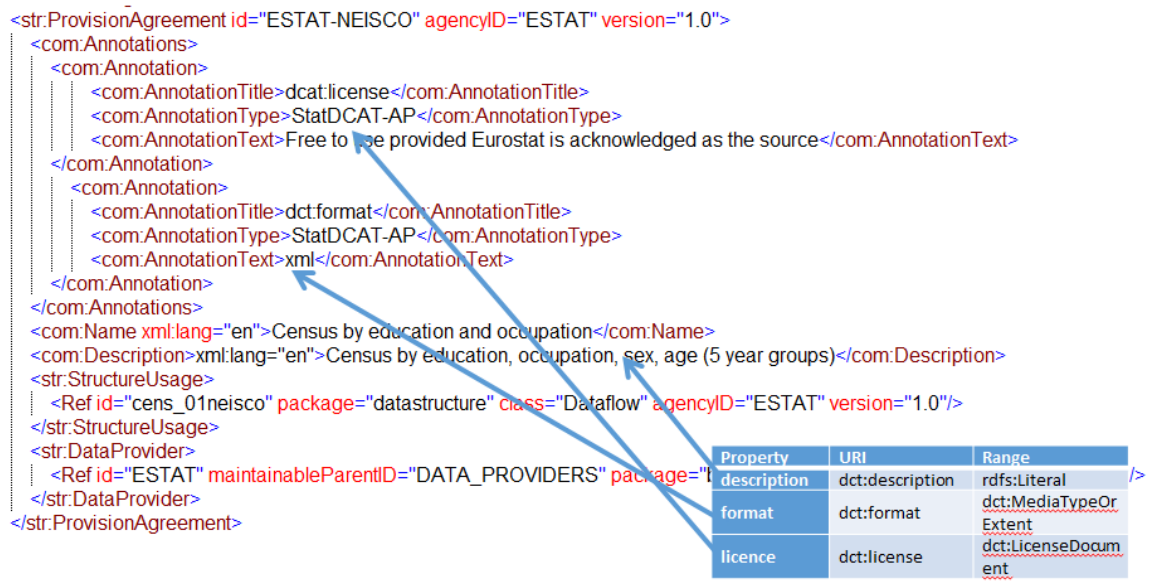


Figure 15: Linking a Distribution to the SDMX Provision Agreement

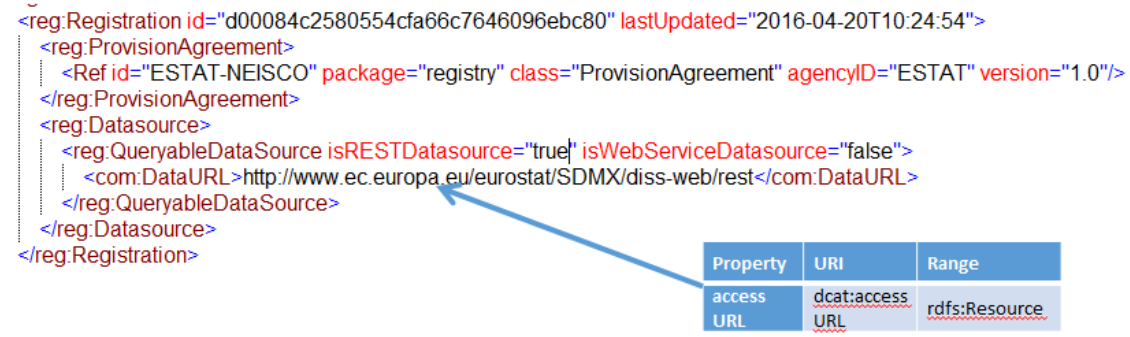


Figure 16: Linking a Distribution (accessURL) to the SDMX Provision Agreement

9.3.7 Agent

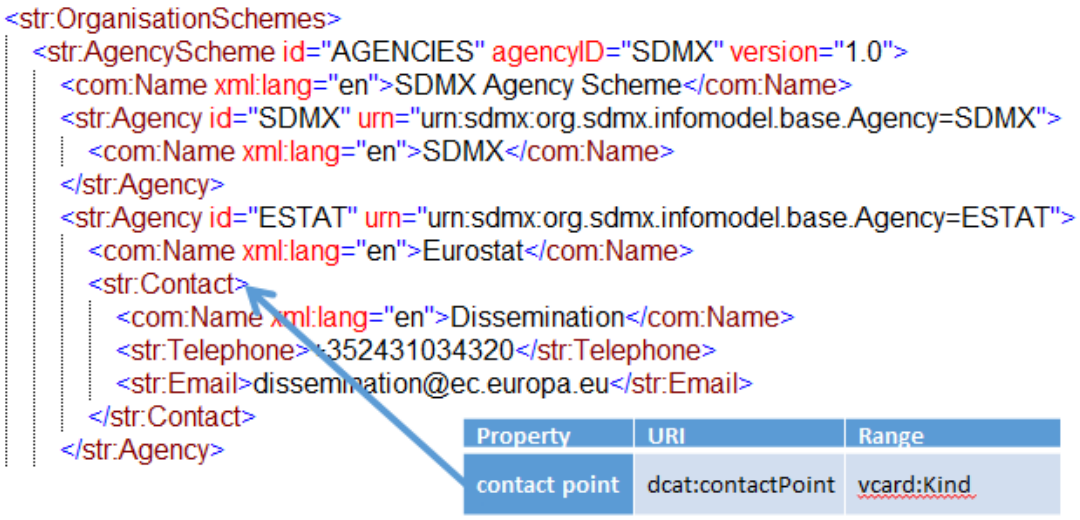


Figure 17:Linking an Agent to the SDMX Agency

9.4 Summary

The mapping above is the recommended mapping between SDMX classes and attributes and DCAT-AP classes and properties. Clearly, an organisation is free to use whatever input source(s) it wishes, including a mixture of sources. The use of SDMX Annotations to curate the DCAT-AP properties is a recommendation for those organisations that wish to use 100% SDMX structural metadata for this mapping. In order to achieve interoperability between systems, StatDCAT-AP will specify a controlled vocabulary for the AnnotationTitle (DCAT-AP property).

10 SDMX-BASED TRANSFORMATION MECHANISM

10.1 Scope of this section

The scope of this section is to describe a mechanism that is intended to assist statistical organisations to create StatDCAT-AP without the need for the organisation to understand the syntax and rules of DCAT-AP. In this document this is referred to as the "Transformation Mechanism".

Whilst any organisation is free to choose whichever mechanism it prefers in order to create and publish DCAT-AP RDF, the Transformation Mechanism described here will be provided in the form of tools that an organisation can use to convert an XML file based on SDMX-formatted structures (SDMX-ML) to DCAT-AP.

The intent of this Transformation Mechanism is to assist those organisations that do not wish to invest in resources to understand RDF technologies and vocabularies and thus to encourage organisations to use DCAT-AP to publish the content of their open data. Whilst the two formats used in this Transformation Mechanism will be familiar to an organisation already using SDMX, the Metadata Set variant of the format is a very simple XML structure and should be easy for an organisation with general XML skills to create the metadata required from its own metadata sources, even if that organisation does not use SDMX.

The Transformation Mechanism is first explained. This is followed by an example of the mapping of the input format used by the Transformation Mechanism to the DCAT-AP properties.

10.2 Transformation mechanism

The essence of this mechanism is shown in the following diagram and explanation.

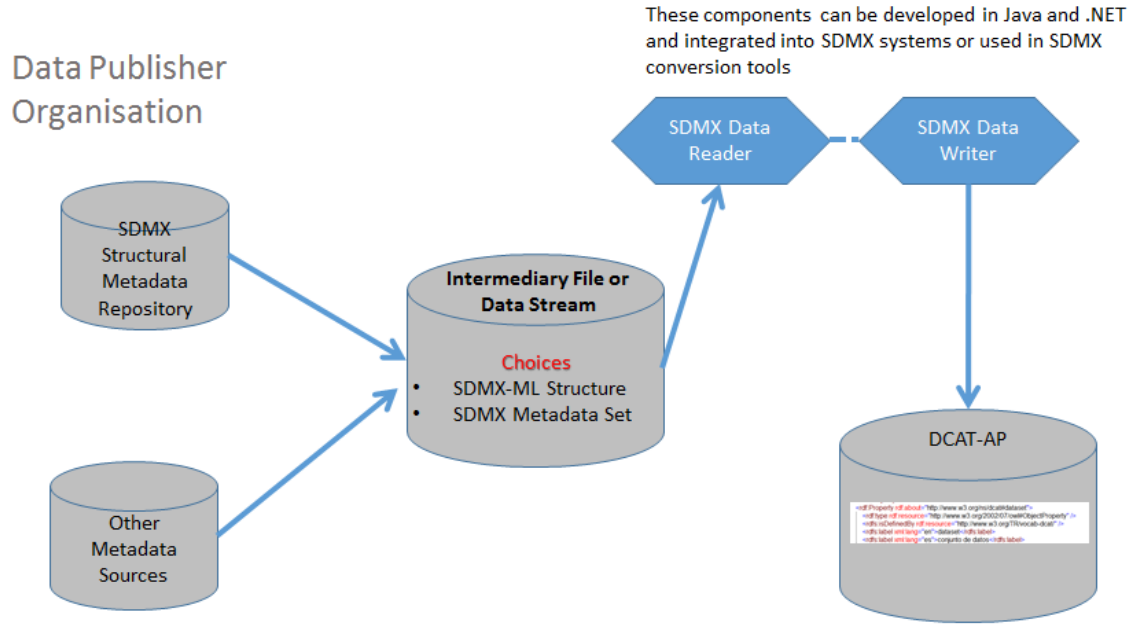


Figure 18: Diagram of the flow of metadata though the Intermediary Mechanism

The structural metadata required to populate the DCAT-AP can be derived from many types of source. The sources may be multiple and may include a maintained structural metadata repository which could be an SDMX-compliant source such as an SDMX Registry.

The metadata required for the intermediary format may be made available either as SDMX structural metadata or as an SDMX metadata set. Both of these options are described later in this section.

The metadata provided is read by a "Data Reader" which understands the format of the metadata stream (i.e. SDMX structural metadata or a SDMX metadata set), and makes these metadata available to a Data Writer via an API that is conformant to the SDMX Common Component Architecture. The Data Writer creates the DCAT-AP output. Therefore, the Transformation Mechanism comprises two Data Readers (one for each of the two formats) and one for the Data Writer. Note that using the SDMX Common Component Architecture the Data Reader and Data Writer have no knowledge of each other and so any Data Reader can supply data to any Data Writer. Thus the Data Readers and Writers can be integrated into an organisation's system or can be built easily into transformation tools. There are a number of SDMX validation and transformation tools that can be extended to use these two Data Readers and the DCAT-AP Data Writer.

10.3 Transformation input formats

10.3.1 Choice of mechanisms

It is the responsibility of the user system to extract the metadata from the metadata source(s) and write the metadata to the relevant transformation input format. So, the question that requires an answer is "why, then, not just create DCAT-AP directly".

The answer is that if the organisation is comfortable with creating DCAT-AP directly from its own systems, then this is the approach it can take. However, if the organisation is not comfortable with this direct approach (e.g. maybe it does not have RDF skills, or it already has SDMX systems in place and is more familiar with SDMX formats) then the Transformation Mechanism is an attractive approach: it uses SDMX formats and has in-built validation to ensure that the metadata are valid for the DCAT-AP.

10.3.2 SDMX Structural Metadata

The format is an SDMX Structure Message. The mapping of SDMX to DCAT-AP has been described in Section 9 of this specification and examples of the mapping are also given in that Section.

However, there is one difference between the mapping given in Section 9 and the format used in the Transformation Mechanism. This concerns the accessURL of the DCAT Distribution. In SDMX the Registration element is not output in the SDMX Structure Message, it is output in the SDMX Registry Interface Message. Therefore, for

the purpose of this transformation this metadata is represented as an Annotation in the Provision Agreement.

So, taking the example from Section 9

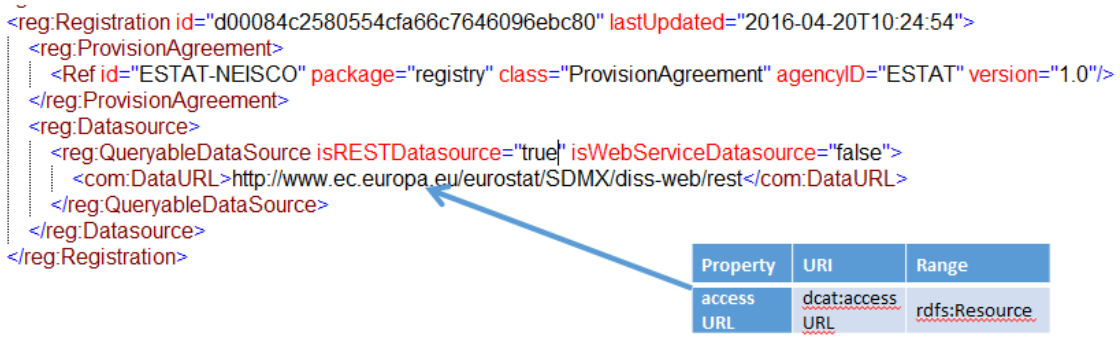


Figure 19: From Section 9 - Linking a Distribution (accessURL) to the SDMX Provision Agreement

Using the Transformation Mechanism this is output as:

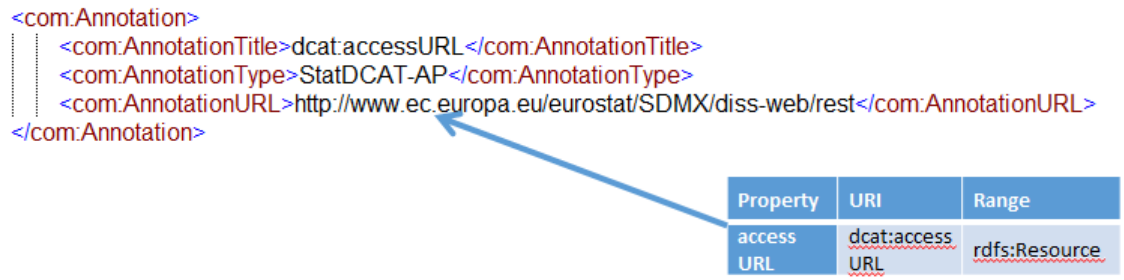


Figure 20: Transformation format - Linking a Distribution (accessURL) to the SDMX Provision Agreement

So, the full example of the Provision Agreement is:

```
<str:ProvisionAgreement id="ESTAT-NEISCO" agencyID="ESTAT" version="1.0">
  <com:Annotations>
    <com:Annotation>
      <com:AnnotationTitle>dcat:license</com:AnnotationTitle>
      <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
      <com:AnnotationText>Free to use provided Eurostat is acknowledged as the source</com:AnnotationText>
    </com:Annotation>
    <com:Annotation>
      <com:AnnotationTitle>dcat:accessURL</com:AnnotationTitle>
      <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
      <com:AnnotationURL>http://www.ec.europa.eu/eurostat/SDMX/diss-web/rest</com:AnnotationURL>
    </com:Annotation>
    <com:Annotation>
      <com:AnnotationTitle>dct:format</com:AnnotationTitle>
      <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
      <com:AnnotationText>xml</com:AnnotationText>
    </com:Annotation>
  </com:Annotations>
  <com:Name xml:lang="en">Census by education and occupation</com:Name>
  <com:Description>xml:lang="en">Census by education, occupation, sex, age (5 year groups)</com:Description>
  <str:StructureUsage>
    <Ref id="cens_01neisco" package="datastructure" class="Dataflow" agencyID="ESTAT" version="1.0"/>
  </str:StructureUsage>
  <str:DataProvider>
    <Ref id="ESTAT" maintainableParentID="DATA_PROVIDERS" package="base" class="DataProvider" agencyID="ESTAT"/>
  </str:DataProvider>
</str:ProvisionAgreement>
```

Figure 21: Example Provision Agreement for DCAT-AP Distribution

The full example Structure Message is shown in Annex I.

It is a simple software development to create an extract process from an SDMX Registry to create the SDMX format required for the Transformation Mechanism, as the relevant metadata can be retrieved using SDMX web services which are supported already by an SDMX Registry.

10.3.3 SDMX Metadata Set

Structure

A Metadata Set represents metadata for some or all of the DCAT Classes and Properties as Metadata Attributes. The structure of a Metadata Set is defined by a Metadata Structure Definition (MSD). The MSD contains all of the information required to structure the content of a Metadata Set in terms of:

For each Metadata Attribute

- The Concept used (i.e. the DCAT-AP Class or Property)
- The valid content (e.g. a Code List, text, URL, integer, no content etc.)
- Child Metadata Attributes if a hierarchy is specified

The MSD also specifies the type of object (class) to which the metadata pertains, such as an SDMX Dataflow. The identification of the actual instance (e.g. an actual Dataflow) is contained in the Metadata Set together with the content of the Metadata Attributes.

A schematic of the MSD is shown below.

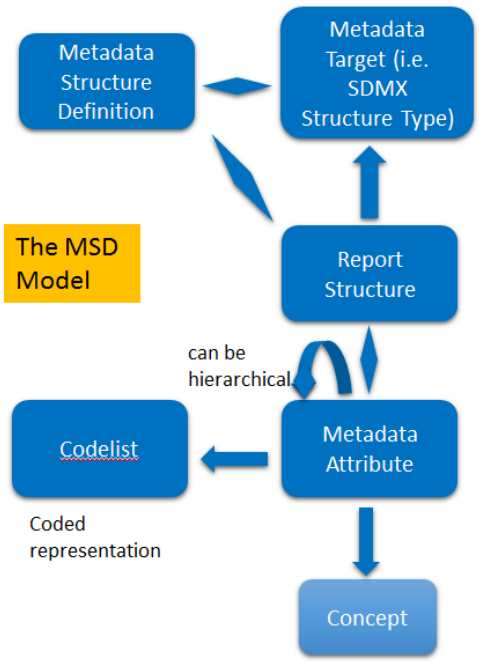


Figure 22: Schematic diagram of the SDMX Metadata Structure Definition

A schematic of the Metadata Set is shown below.

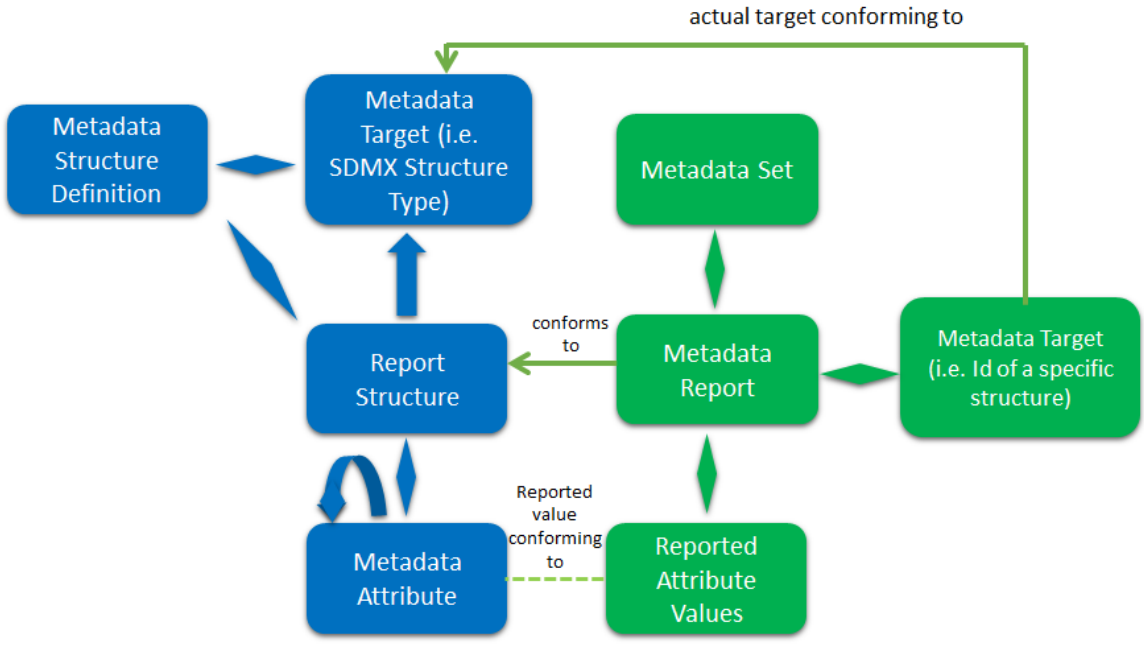


Figure 23: Schematic diagram of the SDMX Metadata Set

The green boxes are the content of the Metadata Set. The MSD is not a part of the Metadata Set but both the MSD and Report Structure are identified in the Metadata Set. The Id of the Metadata Attribute is contained in the Reported Attribute thus enabling the structure and content of the Reported Attribute to be validated. The Metadata Target contains the Id of SDMX structural component to which the metadata

pertains. In the example the target is the SDMX Category Scheme that represents the DCAT-SP Catalogue.

Example MSD

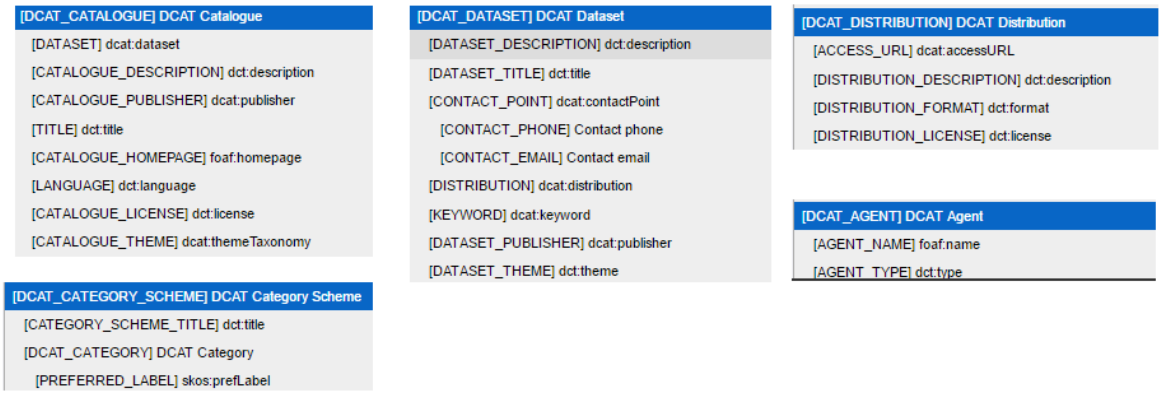


Figure 24: Metadata Attributes in the DCAT-AP MSD

This set of Metadata Attributes represents all of the DCAT-AP mandatory and conditional properties. These are the child attributes of the Metadata Attribute identifying the DCAT-AP class e.g. DCAT_CATALOGUE and its properties. Additional hierarchies are defined where appropriate e.g. in DCAT_DATASET the CONTACT_POINT has two child Metadata Attributes.

The Metadata Attributes representing the DCAT-AP classes are for grouping purposes which enable the transformation software to determine to which DCAT-AP class the metadata pertains. They have no content themselves but have child Metadata Attributes. The following picture shows some examples of the type of valid content that can be specified.

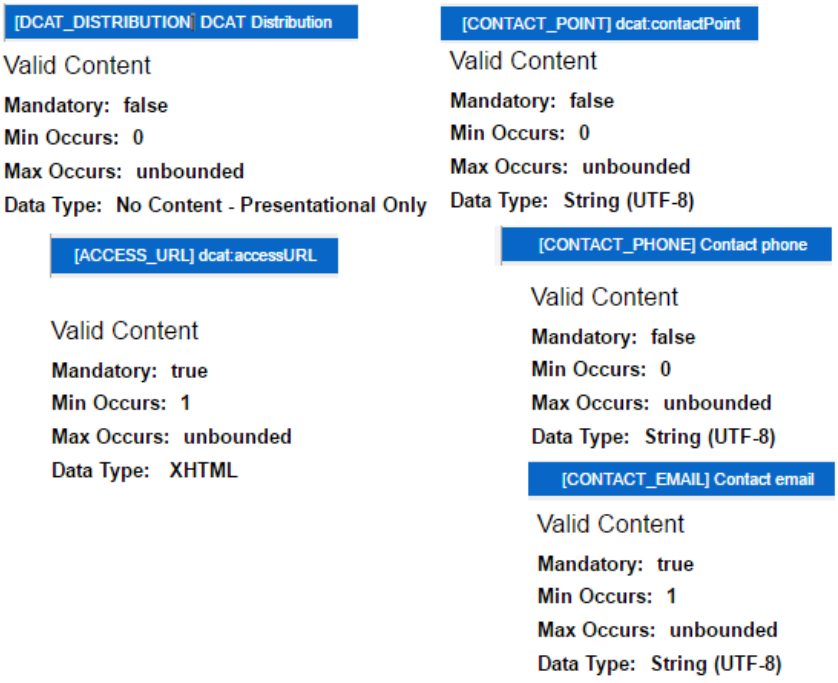


Figure 25: Example Metadata Attribute Specification

The examples above show:

1. DCAT_DISTRIBUTION is grouping only and so no actual value is reported in a Metadata Set.
2. The accessURL is mandatory if the DCAT_DISTRIBUTION is present and its valid representation is XHTML.
3. Contact Point can occur many times and, if present the CONTACT_PHONE is not mandatory but the CONTACT_EMAIL is mandatory.

Note that a code list may be specified as the valid representation, in which case the value of the reported attribute in the Metadata Set must be a code in the assigned code list. There is no example of this in the Figure above.

Example Metadata Report

The following SDMX Metadata Report shows how the DCAT-AP metadata are represented according to the MSD.

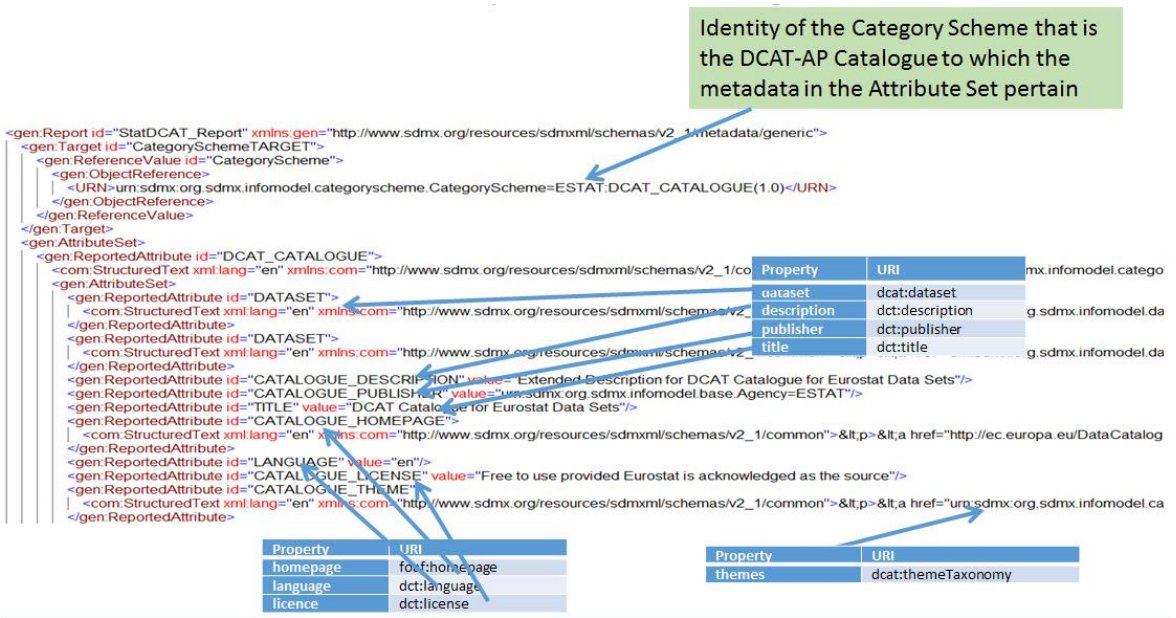


Figure 26: SDMX catalogue metadata pertaining to the DCAT-AP Catalogue

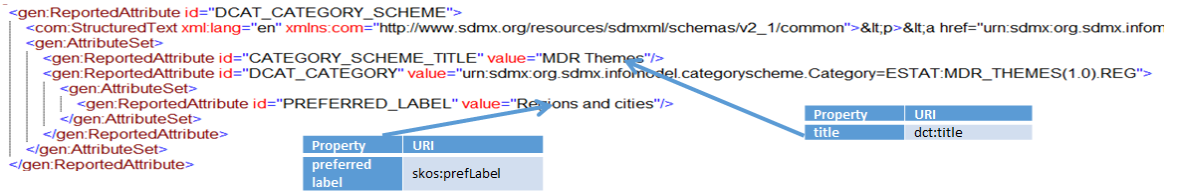


Figure 27: SDMX category scheme metadata pertaining to the DCAT-AP Catalogue

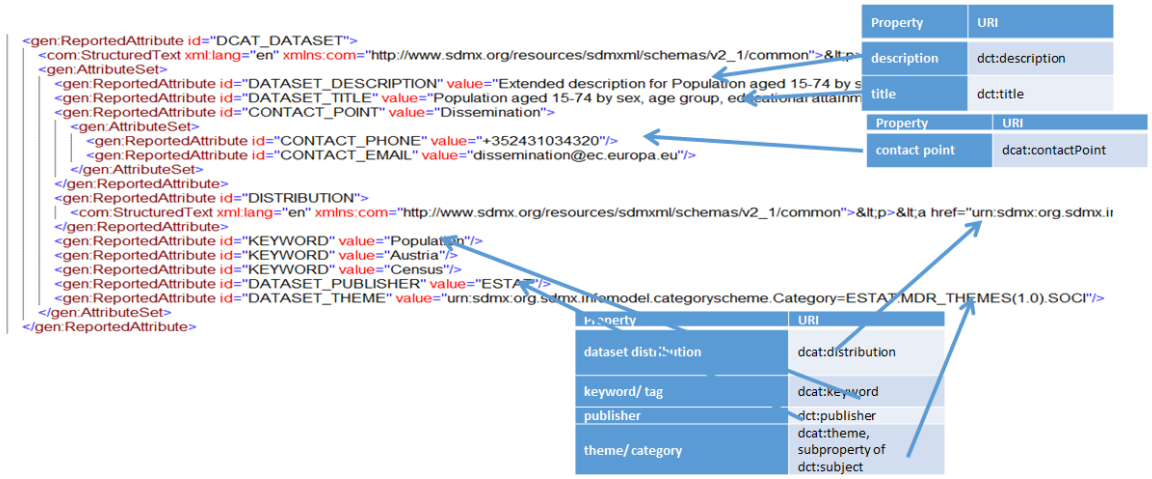


Figure 28: SDMX dataset metadata pertaining to the DCAT-AP Catalogue

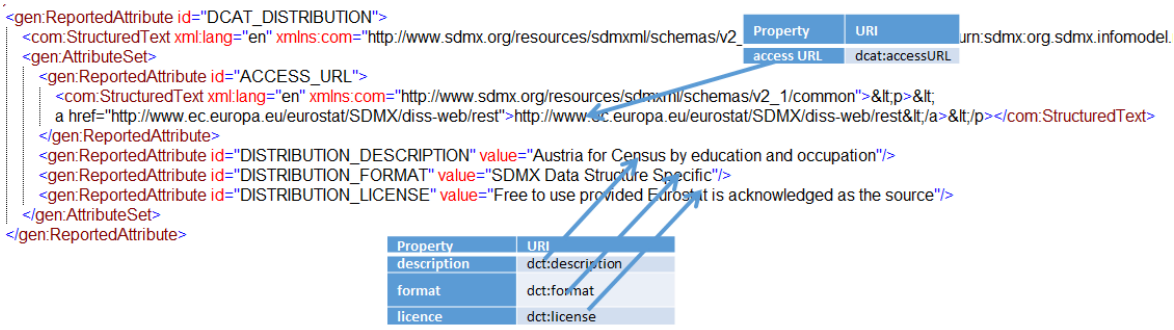


Figure 29: SDMX distribution metadata pertaining to the DCAT-AP Catalogue

It is possible to create a Metadata Set for any or all of the DCAT-AP classes to be supported by StatDCAT-AP. Therefore, an entire catalogue can be published including all the associated Datasets, Distributions, Category Schemes., and Agents. Alternatively, metadata may be added to an existing Catalogue incrementally.

10.4 Advantages and disadvantages of the two transformation formats

10.4.1 SDMX Structure Message

Advantages

- Familiar to organisations using SDMX
- Can be generated easily from an SDMX Registry

Disadvantages

- The XML can be complex and verbose
- Annotations cannot be
    - coded (representation is restricted to text and URL)
    - hierarchical (but there is a mechanism to achieve this)
    - validated by SDMX validators (e.g. that the Title is valid)

    - given mandatory and optional status (all Annotations are optional)
- Could create unnecessary "noise" when exchanging structural metadata with other organisations if this is the source of the metadata in an SDMX Registry-compliant metadata source

However

It would be possible to use the MSD for the Metadata Set option to validate that the content of the structural metadata is complete and that the Annotation metadata is correct (e.g. text representing a coded value can be validated with a code list) and that the correct hierarchy is built in DCAT-AP.

10.4.2 SDMX Metadata Set

Advantages

- Simple XML structure
- Attributes can be:
    - assigned any type of representation (e.g. coded, text, HTML, Boolean etc.)
    - hierarchical
    - validated
    - usage status can be mandatory or optional
- The Attribute Set can reference any object that can be identified (e.g. Dataflow, Provision Agreement, Category Scheme)
- Is separate from the structural metadata so does not affect the structural metadata components
- If present, a Metadata Attribute can be "presentational", just giving structure to child attributes

Disadvantages

- Not always well understood by SDMX users (may result in some reluctance to use this mechanism)
- Not widely used

10.5 Summary

Whilst an organisation can choose to create DCAT-AP directly from its own systems, having an intermediary Transformation Mechanism will be of benefit to some organisations. This will be particularly true for those organisations already using SDMX.

There is a need for all organisations to validate the metadata to ensure that it is compliant with the DCAT-AP classes and properties. The MSD can play a role in the validation regardless of the of the intermediary transformation format because the MSD describes the valid content of DCAT-AP metadata.

The Metadata Set intermediary format is simpler than the SDMX structural metadata. However, for those organisations using an SDMX Registry, it is probable that these registry systems will be able to harvest the metadata and export as DCAT-AP using the Transformation Mechanism.

11 ACKNOWLEDGEMENTS

Table with all participants with affiliation.

Annex I SDMX FILES USED FOR THE EXAMPLES

I.1 SDMX Structural Metadata

```
<?xml version="1.0" encoding="UTF-8"?>
<mes:Structure xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xml="http://www.w3.org/XML/1998/namespace"
xmlns:mes="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/message"
xmlns:str="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/structure"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"
xsi:schemaLocation="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/message
https://registry.sdmx.org/schemas/v2_1/SDMXMessage.xsd">
          <mes:Header>
    <mes:ID>IDREF169</mes:ID>
    <mes:Test>false</mes:Test>
    <mes:Prepared>2016-05-05T15:11:56</mes:Prepared>
    <mes:Sender id="FR"/>
    <mes:Receiver id="not_supplied"/>
  </mes:Header>
  <mes:Structures>
    <str:OrganisationSchemes>
```

DCAT Agent

```
      <str:AgencyScheme id="AGENCIES"
urn="urn:sdmx:org.sdmx.infomodel.base.AgencyScheme=SDMX:AGENCIES(1.0)" isExternalReference="false"
agencyID="SDMX" isFinal="false" version="1.0">
        <com:Name xml:lang="en">SDMX Agency Scheme</com:Name>
          <str:Agency id="ESTAT" urn="urn:sdmx:org.sdmx.infomodel.base.Agency=ESTAT">
          <com:Name xml:lang="en">Eurostat</com:Name>
          <str:Contact>
            <com:Name xml:lang="en">Dissemination</com:Name>
            <str:Telephone>+352431034320</str:Telephone>
            <str:Email>dissemination@ec.europa.eu</str:Email>
          </str:Contact>
        </str:Agency>
      </str:AgencyScheme>
    </str:OrganisationSchemes>
    <str:Dataflows>
```

DCAT Dataset

```
      <str:Dataflow id="DF_HC58" urn="urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:DF_HC58(1.0)"
isExternalReference="false" agencyID="ESTAT" isFinal="false" version="1.0">
        <com:Name xml:lang="en">Hypercube 58</com:Name>
        <str:Structure>
          <Ref id="HC58_DSD" package="datastructure" class="DataStructure" agencyID="ESTAT" version="1.0"/>
        </str:Structure>
      </str:Dataflow>
      <str:Dataflow id="cens_01neisco"
urn="urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:cens_01neisco(1.0)" isExternalReference="false"
agencyID="ESTAT" isFinal="false" version="1.0">
      <com:Annotations>
          <com:Annotation>
            <com:AnnotationTitle>dcat:distribution</com:AnnotationTitle>
            <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
            <com:AnnotationURL>urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ESTAT:AT-
NEISCO(1.0)</com:AnnotationURL>
          </com:Annotation>
          <com:Annotation>
            <com:AnnotationTitle>dcat:keyword</com:AnnotationTitle>
            <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
            <com:AnnotationText>Population</com:AnnotationText>
          </com:Annotation>
          <com:Annotation>
            <com:AnnotationTitle>dcat:keyword</com:AnnotationTitle>
            <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
            <com:AnnotationText>Austria</com:AnnotationText>
          </com:Annotation>
          <com:Annotation>
            <com:AnnotationTitle>dcat:keyword</com:AnnotationTitle>
            <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
            <com:AnnotationText>Census</com:AnnotationText>
          </com:Annotation>
```

```
            <com:Annotation>
              <com:AnnotationTitle>dcat:theme</com:AnnotationTitle>
              <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
  <com:AnnotationURL>urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).SOC</com:
AnnotationURL>
            </com:Annotation>
          </com:Annotations>
        <com:Name xml:lang="de">Bevölkerung im Alter zwischen 15 und 74 Jahren nach Geschlecht, Altersklasse,
erreichtes Bildungsniveau (ISCED 1997) und Beruf (ISCO-88)</com:Name>
        <com:Name xml:lang="fr">Population âgée de 15 à 74 ans, par sexe, groupe d'âge, niveau d'instruction (ISCED
1997) et profession (CITP-88)</com:Name>
        <com:Name xml:lang="en">Population by education and occupation</com:Name>
        <com:Description xml:lang="en">Population aged 15-74 by sex, age group, educational attainment (ISCED 1997)
and occupation (ISCO 1988)</com:Description>
        <str:Structure>
          <Ref id="CENS_01_NEISCO" package="datastructure" class="DataStructure" agencyID="ESTAT"
version="1.0"/>
        </str:Structure>
      </str:Dataflow>
    </str:Dataflows>
    <str:CategorySchemes>
```


DCAT Catalogue


```
      <str:CategoryScheme id="DCAT_CATALOGUE"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:DCAT_CATALOGUE(1.0)"
isExternalReference="false" agencyID="ESTAT" isFinal="false" version="1.0">
        <com:Annotations>
          <com:Annotation>
            <com:AnnotationTitle>dcat:dataset</com:AnnotationTitle>
            <com:AnnotationType>StatDCAT-AP</com:AnnotationType>

  <com:AnnotationURL>urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:DF_HC58(1.0)</com:AnnotationUR
L>
          </com:Annotation>
          <com:Annotation>
            <com:AnnotationTitle>foaf:homepage</com:AnnotationTitle>
            <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
            <com:AnnotationURL>http://ec.europa.eu/DataCatalogue</com:AnnotationURL>
          </com:Annotation>
          <com:Annotation>
            <com:AnnotationTitle>dct:license</com:AnnotationTitle>
            <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
            <com:AnnotationText xml:lang="en">Free to use provided Eurostat is acknowledged as the
source</com:AnnotationText>
          </com:Annotation>
        </com:Annotations>
        <com:Name xml:lang="en">DCAT Catalogue for Eurostat Data Sets</com:Name>
        <str:Category id="DATASETS"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:DCAT_CATALOGUE(1.0).DATASETS">
          <com:Name xml:lang="en">Links to Data Sets (Dataflows)</com:Name>
        </str:Category>
        <str:Category id="TOPIC_THEMES"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:DCAT_CATALOGUE(1.0).TOPIC_THEMES">
          <com:Name xml:lang="en">Links to Topic (Category) Schemes</com:Name>
        </str:Category>
      </str:CategoryScheme>
```


DCAT Category Scheme


```
      <str:CategoryScheme id="MDR_THEMES"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR_THEMES(1.0)"
isExternalReference="false" agencyID="ESTAT" isFinal="false" version="1.0">
        <com:Name xml:lang="en">MDR Themes</com:Name>
        <str:Category id="AGRI"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).AGRI">
          <com:Name xml:lang="en">Agriculture, fisheries, forestry and food</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering such domains as agriculture, fisheries,
forestry or food.</com:Description>
        </str:Category>
        <str:Category id="ECON"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).ECON">
          <com:Name xml:lang="en">Economy and finance</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering such domains as economy or
finance.</com:Description>
        </str:Category>
```

```
        <str:Category id="EDUC"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).EDUC">
          <com:Name xml:lang="en">Education, culture and sport</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering such domains as education, culture or
sport.</com:Description>
        </str:Category>
        <str:Category id="ENER"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).ENER">
          <com:Name xml:lang="en">Energy</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering the domain of
energy.</com:Description>
        </str:Category>
        <str:Category id="ENVI"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).ENVI">
          <com:Name xml:lang="en">Environment</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering the domain of
environment</com:Description>
        </str:Category>
        <str:Category id="GOVE"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).GOVE">
          <com:Name xml:lang="en">Government and public sector</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering such domains as government or public
sector.</com:Description>
        </str:Category>
        <str:Category id="HEAL"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).HEAL">
          <com:Name xml:lang="en">Health</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering the domain of
health.</com:Description>
        </str:Category>
        <str:Category id="INTR"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).INTR">
          <com:Name xml:lang="en">International issues</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering the domain of international
issues.</com:Description>
        </str:Category>
        <str:Category id="JUST"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).JUST">
          <com:Name xml:lang="en">Justice, legal system and public safety</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering such domains as justice, legal system
or public safety.</com:Description>
        </str:Category>
        <str:Category id="REGI"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).REGI">
          <com:Name xml:lang="en">Regions and cities</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering such domains as regions or
cities.</com:Description>
        </str:Category>
        <str:Category id="SOCI"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).SOCI">
          <com:Name xml:lang="en">Population and society</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering such domains as population or
society.</com:Description>
        </str:Category>
        <str:Category id="TECH"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).TECH">
          <com:Name xml:lang="en">Science and technology</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering such domains as science or
technology.</com:Description>
        </str:Category>
        <str:Category id="TRAN"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).TRAN">
          <com:Name xml:lang="en">Transport</com:Name>
          <com:Description xml:lang="en">This concept identifies datasets covering such domains as
transport</com:Description>
        </str:Category>
      </str:CategoryScheme>
    </str:CategorySchemes>
    <str:Categorisations>
```

Link between Dataflow and the Category in the MDR Scheme of Topics

```
      <str:Categorisation id="4880e39f-585a-4452-2403-4ea6806df530"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Categorisation=ESTAT:4880e39f-585a-4452-2403-
4ea6806df530(1.0)" isExternalReference="false" agencyID="ESTAT" isFinal="false" version="1.0">
```

```
        <com:Name xml:lang="en">4880e39f-585a-4452-2403-4ea6806df530</com:Name>
        <str:Source>
          <Ref id="cens_01neisco" package="datastructure" class="Dataflow" agencyID="ESTAT" version="1.0"/>
        </str:Source>
        <str:Target>
          <Ref id="SOCI" maintainableParentID="MDR_THEMES" package="categoryscheme" class="Category"
agencyID="ESTAT" maintainableParentVersion="1.0"/>
        </str:Target>
      </str:Categorisation>
```

Links between Dataflows and DATASET Category of the DCAT-Catalogue

```
      <str:Categorisation id="5c96f6c8-67a3-805f-7a2b-81bee0e4f6f4"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Categorisation=ESTAT:5c96f6c8-67a3-805f-7a2b-81bee0e4f6f4(1.0)"
isExternalReference="false" agencyID="ESTAT" isFinal="false" version="1.0">
        <com:Name xml:lang="en">5c96f6c8-67a3-805f-7a2b-81bee0e4f6f4</com:Name>
        <str:Source>
          <Ref id="DF_HC58" package="datastructure" class="Dataflow" agencyID="ESTAT" version="1.0"/>
        </str:Source>
        <str:Target>
          <Ref id="DATASETS" maintainableParentID="DCAT_CATALOGUE" package="categoryscheme"
class="Category" agencyID="ESTAT" maintainableParentVersion="1.0"/>
        </str:Target>
      </str:Categorisation>
      <str:Categorisation id="7ca03df9-a610-4cdc-3c1c-f8dbc7913cd4"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Categorisation=ESTAT:7ca03df9-a610-4cdc-3c1c-f8dbc7913cd4(1.0)"
isExternalReference="false" agencyID="ESTAT" isFinal="false" version="1.0">
        <com:Name xml:lang="en">7ca03df9-a610-4cdc-3c1c-f8dbc7913cd4</com:Name>
        <str:Source>
          <Ref id="cens_01neisco" package="datastructure" class="Dataflow" agencyID="ESTAT" version="1.0"/>
        </str:Source>
        <str:Target>
          <Ref id="DATASETS" maintainableParentID="DCAT_CATALOGUE" package="categoryscheme"
class="Category" agencyID="ESTAT" maintainableParentVersion="1.0"/>
        </str:Target>
      </str:Categorisation>
```

Link between DCAT-Catalogue and the Category Scheme of Topics

```
      <str:Categorisation id="LINK_TO_MDR_TOPICS"
urn="urn:sdmx:org.sdmx.infomodel.categoryscheme.Categorisation=ESTAT:LINK_TO_MDR_TOPICS(1.0)"
isExternalReference="false" agencyID="ESTAT" isFinal="false" version="1.0">
        <com:Name xml:lang="en">Link to MDR Topics</com:Name>
        <str:Source>
          <Ref id="MDR_THEMES" package="categoryscheme" class="CategoryScheme" agencyID="ESTAT"
version="1.0"/>
        </str:Source>
        <str:Target>
          <Ref id="TOPIC_THEMES" maintainableParentID="DCAT_CATALOGUE" package="categoryscheme"
class="Category" agencyID="ESTAT" maintainableParentVersion="1.0"/>
        </str:Target>
      </str:Categorisation>
    </str:Categorisations>
```

DCAT Distribution

```
      <str:ProvisionAgreements>
        <str:ProvisionAgreement id="ESTAT-NEISCO"
urn="urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ESTAT:ESTAT-NEISCO(1.0)"
isExternalReference="false" agencyID="ESTAT" isFinal="false" version="1.0">
        <com:Annotations>
          <com:Annotation>
            <com:AnnotationTitle>dcat:license</com:AnnotationTitle>
            <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
            <com:AnnotationText>Free to use provided Eurostat is acknowledged as the source</com:AnnotationText>
          </com:Annotation>
          <com:Annotation>
            <com:AnnotationTitle>dcat:accessURL</com:AnnotationTitle>
            <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
            <com:AnnotationURL>http://www.ec.europa.eu/eurostat/SDMX/diss-web/rest/</com:AnnotationURL>
          </com:Annotation>
        </com:Annotations>
        <com:Name xml:lang="en">Census by education and occupation</com:Name>
          <com:Description>xml:lang="en">Census by education and occupation,sex, age (5-year
groups)</com:Description>
        <str:StructureUsage>
          <Ref id="cens_01neisco" package="datastructure" class="Dataflow" agencyID="ESTAT" version="1.0"/>
        </str:StructureUsage>
```

```
            <str:DataProvider>
              <Ref id="ESTAT" maintainableParentID="DATA_PROVIDERS" package="base" class="DataProvider"
agencyID="ESTAT" maintainableParentVersion="1.0"/>
            </str:DataProvider>
          </str:ProvisionAgreement>
        </str:ProvisionAgreements>
      </mes:Structures>
    </mes:Structure>
```

I.2 SDMX Metadata Set

I.2.1 Content

The figures below show the content of the Metadata Set, first as viewed in a simple rendering of the content, followed by the SDMX-ML of that portion of the Metadata Set.

Metadata Set – Start

```
<mes:MetadataSet structureRef="MDS2" setID="ba70fc24-f95b-4f0f-a2e2-658d993c2078">
  <com:Name xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common">DCAT_CATALOGUE_1</com:
Name>
  <gen:Report id="StatDCAT_Report"
xmlns:gen="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/metadata/generic">
    <gen:Target id="CategorySchemeTARGET">
      <gen:ReferenceValue id="CategoryScheme">
        <gen:ObjectReference>
<URN>urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:DCAT_CATALOGUE(1.0)</
URN>
        </gen:ObjectReference>
      </gen:ReferenceValue>
    </gen:Target>
    <gen:AttributeSet>
```

Followed by the Reported Attributes for the properties of the various DCAT-AP classes

DCAT Catalogue
urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:DCAT_CATALOGUE(1.0)

dcat:dataset
urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:cens_01neisco(1.0

dcat:dataset
urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:DF_HC58(1.0)

dct:description
Extended Description for DCAT Catalogue for Eurostat Data Sets
dcat:publisher
urn:sdmx:org.sdmx.infomodel.base.Agency=ESTAT
dct:title
DCAT Catalogue for Eurostat Data Sets
foaf:homepage
http://ec.europa.eu/DataCatalogue

dct:language
en
dct:license
Free to use provided Eurostat is acknowledged as the source
dcat:themeTaxonomy
urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR_THEMES(1.0)

<gen:ReportedAttribute id="DCAT_CATALOGUE">
<com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"><p><a
href="urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:DCAT_CATALOGUE(1.0)">u
rn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:DCAT_CATALOGUE(1.0)</
p></com:StructuredText>
        <gen:AttributeSet>
    <gen:ReportedAttribute id="DATASET">
    <com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"><p><a
href="urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:cens_01neisco(1.0">urn:sdmx:org.sdmx.i
nfomodel.datastructure.Dataflow=ESTAT:cens_01neisco(1.0</p></com:StructuredText>
    </gen:ReportedAttribute>
    <gen:ReportedAttribute id="DATASET">
    <com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"><p><a
href="urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:DF_HC58(1.0)">urn:sdmx:org.sdmx.info
model.datastructure.Dataflow=ESTAT:DF_HC58(1.0)</p></com:StructuredText>
    </gen:ReportedAttribute>
    <gen:ReportedAttribute id="CATALOGUE_DESCRIPTION" value="Extended Description for DCAT
Catalogue for Eurostat Data Sets"/>
    <gen:ReportedAttribute id="CATALOGUE_PUBLISHER"
value="urn:sdmx:org.sdmx.infomodel.base.Agency=ESTAT"/>
    <gen:ReportedAttribute id="TITLE" value="DCAT Catalogue for Eurostat Data Sets"/>
    <gen:ReportedAttribute id="CATALOGUE_HOMEPAGE">
    <com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"><p><a
href="http://ec.europa.eu/DataCatalogue">http://ec.europa.eu/DataCatalogue</p></com:StructuredT
ext>
    </gen:ReportedAttribute>
    <gen:ReportedAttribute id="LANGUAGE" value="en"/>
    <gen:ReportedAttribute id="CATALOGUE_LICENSE" value="Free to use provided Eurostat is
acknowledged as the source"/>
    <gen:ReportedAttribute id="CATALOGUE_THEME">
    <com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"><p><a
href="urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR_THEMES(1.0)">urn:sd
mx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR_THEMES(1.0)</p></com:S
tructuredText>
    </gen:ReportedAttribute>
  </gen:AttributeSet>
</gen:ReportedAttribute>

DCAT Category Scheme
urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR_THEMES(1.0)

dct:title
MDR Themes
DCAT Category
urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).AGRI
skos:prefLabel
Agriculture, fisheries, forestry and food
DCAT Category
urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).ECON
skos:prefLabel
Economy and finance
DCAT Category
urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).EDUC
skos:prefLabel
Education, culture and sport
DCAT Category
urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).ENER
skos:prefLabel
Energy

And so on….

```
<gen:ReportedAttribute id="DCAT_CATEGORY_SCHEME">
          <com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common">&lt;p&gt;&lt;a
href="urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR_THEMES(1.0)">urn:sd
mx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR_THEMES(1.0)&lt;/a&gt;&lt;/p></com:S
tructuredText>
  <gen:AttributeSet>
    <gen:ReportedAttribute id="CATEGORY_SCHEME_TITLE" value="MDR Themes"/>
    <gen:ReportedAttribute id="DCAT_CATEGORY"
value="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).AGRI">
      <gen:AttributeSet>
      <gen:ReportedAttribute id="PREFERRED_LABEL" value="Agriculture, fisheries, forestry and food"/>
      </gen:AttributeSet>
    </gen:ReportedAttribute>
    gen:ReportedAttribute id="DCAT_CATEGORY"
value="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).ECON">
      <gen:AttributeSet>
      <gen:ReportedAttribute id="PREFERRED_LABEL" value="Economy and finance"/>
      </gen:AttributeSet>
    </gen:ReportedAttribute>
    <gen:ReportedAttribute id="DCAT_CATEGORY"
value="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).EDUC">
      <gen:AttributeSet>
      <gen:ReportedAttribute id="PREFERRED_LABEL" value="Education, culture and sport"/>
      </gen:AttributeSet>
    </gen:ReportedAttribute>
    <gen:ReportedAttribute id="DCAT_CATEGORY"
value="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).ENER">
      <gen:AttributeSet>
      <gen:ReportedAttribute id="PREFERRED_LABEL" value="Energy"/>
      </gen:AttributeSet>
    </gen:ReportedAttribute>
```

And so on….

DCAT Dataset
urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:cens_01neisco(1.0)

dct:description
Extended description for Population aged 15-74 by sex, age group, educational attainment (ISCED 1997) and occupation (ISCO 1988)
dct:title
Population aged 15-74 by sex, age group, educational attainment (ISCED 1997) and occupation (ISCO 1988)
dcat:contactPoint
Dissemination
Contact phone
+352431034320
Contact email
dissemination@ec.europa.eu
dcat:distribution
urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ESTAT:AT-NEISCO(1.0)

dcat:keyword
Population
dcat:keyword
Austria
dcat:keyword
Census
dcat:publisher
ESTAT
dct:theme
urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).SOCI

```
<gen:ReportedAttribute id="DCAT_DATASET">
  <com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common">&lt;p&gt;&lt;a
href="urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:cens_01neisco(1.0)">urn:sdmx:org.sdmx.
infomodel.datastructure.Dataflow=ESTAT:cens_01neisco(1.0)&lt;/a&gt;&lt;/p></com:StructuredText>
  <gen:AttributeSet>
    <gen:ReportedAttribute id="DATASET_DESCRIPTION" value="Extended description for Population aged
15-74 by sex, age group, educational attainment (ISCED 1997) and occupation (ISCO 1988)"/>
    <gen:ReportedAttribute id="DATASET_TITLE" value="Population aged 15-74 by sex, age group,
educational attainment (ISCED 1997) and occupation (ISCO 1988)"/>
    <gen:ReportedAttribute id="CONTACT_POINT" value="Dissemination">
      <gen:AttributeSet>
      <gen:ReportedAttribute id="CONTACT_PHONE" value="+352431034320"/>
      <gen:ReportedAttribute id="CONTACT_EMAIL" value="dissemination@ec.europa.eu"/>
      </gen:AttributeSet>
    </gen:ReportedAttribute>
    <gen:ReportedAttribute id="DISTRIBUTION">
      <com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common">&lt;p&gt;&lt;a
href="urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ESTAT:AT-
NEISCO(1.0)">urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ESTAT:AT-
NEISCO(1.0)&lt;/a&gt;&lt;/p></com:StructuredText>
    </gen:ReportedAttribute>
    <gen:ReportedAttribute id="KEYWORD" value="Population"/>
    <gen:ReportedAttribute id="KEYWORD" value="Austria"/>
    <gen:ReportedAttribute id="KEYWORD" value="Census"/>
    <gen:ReportedAttribute id="DATASET_PUBLISHER" value="ESTAT"/>
    <gen:ReportedAttribute id="DATASET_THEME"
value="urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).SOCI"/>
  </gen:AttributeSet>
</gen:ReportedAttribute>
```

DCAT Distribution
urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ESTAT:AT-NEISCO(1.0)

dcat:accessURL
http://localhost:8080/FusionRegistry/ws/rest

dct:description
Austria for Census by education and occupation
dct:format
SDMX Data Structure Specific
dct:license
Free to use provided Eurostat is acknowledged as the source

<gen:ReportedAttribute id="DCAT_DISTRIBUTION">
<com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"><p><a
href="urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ESTAT:AT-
NEISCO(1.0)">urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ESTAT:AT-
NEISCO(1.0)</p></com:StructuredText>
<gen:AttributeSet>
<gen:ReportedAttribute id="ACCESS_URL">
<com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"><p><a
href="http://localhost:8080/FusionRegistry/ws/rest">http://localhost:8080/FusionRegistry/ws/rest</p><
/com:StructuredText>
</gen:ReportedAttribute>
<gen:ReportedAttribute id="DISTRIBUTION_DESCRIPTION" value="Austria for Census by education and
occupation"/>
<gen:ReportedAttribute id="DISTRIBUTION_FORMAT" value="SDMX Data Structure Specific"/>
<gen:ReportedAttribute id="DISTRIBUTION_LICENSE" value="Free to use provided Eurostat is
acknowledged as the source"/>
</gen:AttributeSet>
</gen:ReportedAttribute>

DCAT Agent
urn:sdmx:org.sdmx.infomodel.base.Agency=ESTAT

foaf:name
Eurostat
dct:type
Publisher

<gen:ReportedAttribute id="DCAT_AGENT">
<com:StructuredText xml:lang="en"
xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"><p><a
href="urn:sdmx:org.sdmx.infomodel.base.Agency=ESTAT">urn:sdmx:org.sdmx.infomodel.base.Agency=EST
AT</p></com:StructuredText>
<gen:AttributeSet>
<gen:ReportedAttribute id="AGENT_NAME" value="Eurostat"/>
<gen:ReportedAttribute id="AGENT_TYPE" value="Publisher"/>
</gen:AttributeSet>
</gen:ReportedAttribute>

End of Metadata Set

</gen:AttributeSet>
</gen:Report>
</mes:MetadataSet>