SC8DI07171

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

Document Metadata

Date	2016-07-04
Status	For public review
Version	0.12
Authors	Makx Dekkers – AMI Consult Chris Nelson – Metadata Technologies Stefanos Kotoglou – PwC EU Services
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:

Table of Contents

List of Figures

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 conceptual mapping of

[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/

SDMX to DCAT-AP was also undertaken both on the metadata level (assessing "reference" metadata created by Eurostat using the ESMS as the standardised structure definition for creating 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 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 ambition 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 outlines the possible approaches towards export of data from existing systems into the StatDCAT-AP.

Section 8 contains a conformance statement.

Section 9 describes the property for relating an Agent to a Dataset.

Section 10 includes clarifications regarding accessibility and multilingual aspects.

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

Annex I provides an overview of all the classes and properties.

Annex II describes the mapping of SDMX to DCAT.

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

Annex IV includes the SDMX files that were used for the examples.

2 TERMINOLOGY USED IN THIS DOCUMENT

Application Profile	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.
Catalogue	A curated collection of metadata about datasets.
Catalogue record	A set of statements about the description of a dataset in the catalogue, e.g. providing information about when a dataset was entered in the catalogue or when its description was modified.
Data Cube Vocabulary	A W3C Recommendation[9] that specifies an RDF vocabulary designed to facilitate publication of multi-dimensional data, such as statistics, on the Web in such a way that it can be linked to related datasets and concepts.
Data Portal	A Web-based system that contains a data catalogue with descriptions of datasets and provides services enabling discovery and re-use of the datasets.
Dataset	A collection of data, published or curated by a single source, and available for access or download in one or more formats.
DCAT – Data Catalog Vocabulary	A W3C Recommendation[10] that specifies an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web.
Distribution	A specific available form of a dataset. If a dataset is published in multiple formats (e.g. Excel, CSV, Data Cube) these are described as separate distributions.
Data Structure Definition	Set of structural metadata associated to a dataset, which includes information about how concepts are associated with the measures, dimensions, and attributes of a data cube, along with information about the representation of data and related descriptive metadata.

[9] W3C. The RDF Data Cube Vocabulary. W3C Recommendation 16 January 2014. https://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/
[10] W3C. Data Catalog Vocabulary (DCAT). W3C Recommendation 16 January 2014. http://www.w3.org/TR/vocab-dcat/

MDR – Metadata Registry	The Metadata Registry[11] registers and maintains definition data (metadata elements, named authority lists, schemas, etc.) used by the different European Institutions involved in the legal decision making process gathered in the Interinstitutional Metadata Maintenance Committee (IMMC) and by the Publications Office of the EU in its production and dissemination process.
Metadata Structure Definition	Specification of the allowed content of a metadata set in terms of attributes for which content is to be provided and to which type of object the metadata pertain.
Reference metadata	Metadata describing the contents and the quality of the statistical data.
SDMX	An International Standard (ISO 17369:2013)[12] that provides an integrated approach to facilitating Statistical Data and Metadata Exchange (SDMX), enabling interoperable implementations within and between systems concerned with the exchange, reporting and dissemination of statistical data and related metadata
SDMX cross-domain concepts	A set of standard concepts, covering structural and reference metadata, which should be used in several statistical domains wherever possible to enhance possibilities of the exchange of data and metadata between organisations.
Structural metadata	Metadata that identify and describe data and reference metadata.
URI – Uniform Resource Identifier	An IETF Request for Comments (RFC)[13] specifying a compact sequence of characters that identifies an abstract or physical resource. URIs on the Web are a subset of URLs and are often called HTTP URIs.
URL – Uniform Resource Locator	An IETF Request for Comments (RFC)[14] specifying the syntax and semantics of formalized information for location and access of resources via the Internet.

[11] Publications Office of the European Union. Metadata Registry (MDR). http://publications.europa.eu/mdr/
[12] Statistical Data and Metadata eXchange. https://sdmx.org/
[13] IETF. RFC 3986. Uniform Resource Identifier (URI): Generic Syntax. https://www.ietf.org/rfc/rfc3986.txt
[14] IETF. RFC 1738. Uniform Resource Locators (URL). https://www.ietf.org/rfc/rfc1738.txt

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)[15], the Publications Office and Eurostat collaborate on the automated harvesting of metadata from Eurostat 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)[16]. 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 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, 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: 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, which is now recognised as ISO International Standard IS-17369, are presented in Figure 1.

[15] European Union Open Data Portal. http://open-data.europa.eu
[16] See the file ESTAT_xxx.zip in http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing

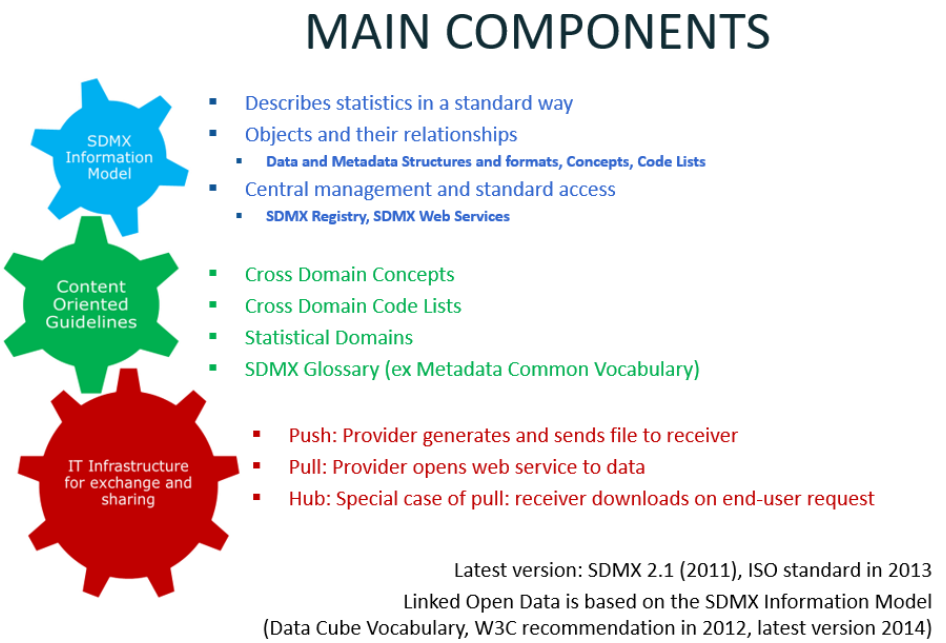


Figure 1: SDMX Main Components

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

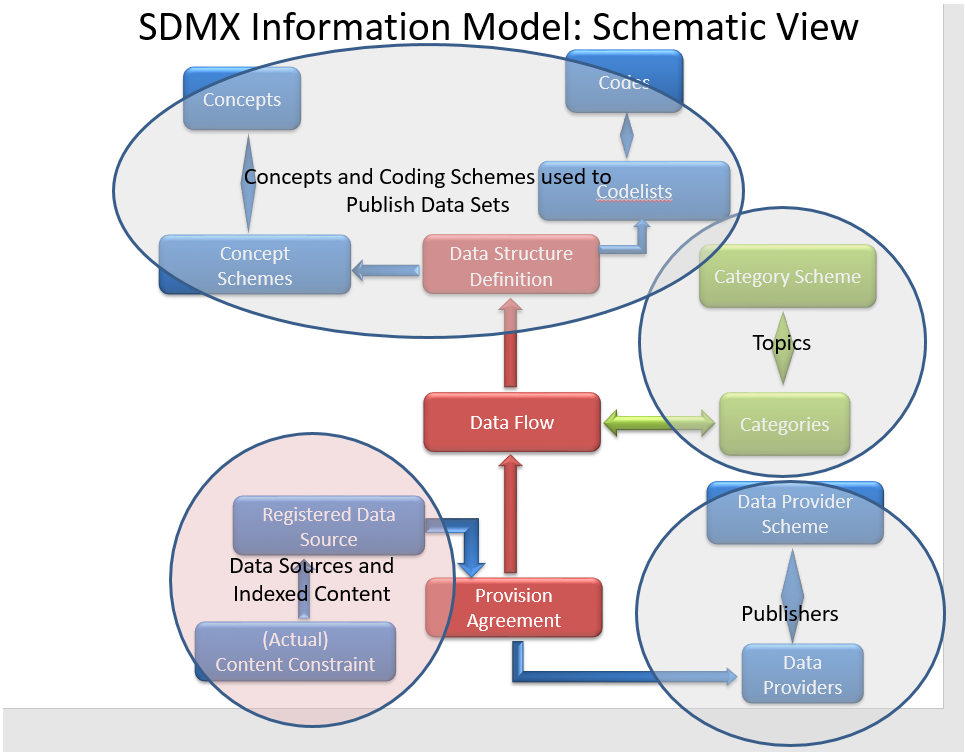


Figure 2: SDMX Information Model: Schematic View

3.1.3 ESMS

The Euro SDMX Metadata Structure (ESMS)[17] 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 (i.e. allowed content) 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 both 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.).

A fragment of the ESMS specification (release 4, 2014) is shown in Figure 3.

	Concept Name	Concept Code	Descriptions	Representation	ESS Guidelines
1	**Contact**	CONTACT	Individual or organisational contact points for the data or metadata, including information on how to reach the contact		
1.1	Contact organisation	CONTACT_ORGANISATION	The name of the organisation of the contact points for the data or metadata.	Text	The full name of your organisation.
1.2	Contact organisation unit	ORGANISATION_UNIT	An addressable subdivision of an organisation.	Text	The name of the unit responsible for the metadata file (it can also include a unit number).
1.3	Contact name	CONTACT_NAME	The name of the contact points for the data or metadata.	Text	The name of the person responsible for the statistical domain (first name and family name).
1.4	Contact person function	CONTACT_FUNCT	The area of technical responsibility of the contact, such as "methodology", "database management" or	Text	The title of the person responsible for the statistical domain (this title can contain the precise area of responsibility such as methodologist or data base manager)
1.5	Contact mail address	CONTACT_MAIL	The postal address of the contact points for the data or metadata.	Text	The postal address of the person responsible for the statistical domain.
1.6	Contact email address	CONTACT_EMAIL	E-mail address of the contact points for the data or metadata.	e-mail	The email address of the person responsible for the statistical domain (this can be an individual mail address or a functional
1.7	Contact phone number	CONTACT_PHONE	The telephone number of the contact points for the data or metadata.	Telephone	The phone number of the person responsible for the statistical domain.
1.8	Contact fax number	CONTACT_FAX	Fax number of the contact points for the data or metadata.	Fax	The fax number of the person responsible for the statistical domain.
2	**Metadata update**	META_UPDATE	The date on which the metadata element was inserted or modified in the database.		
2.1	Metadata last certified	META_CERTIFIED	Date of the latest certification provided by the domain manager to confirm that the metadata posted are still up-to-date, even if the content has not been amended.	Date	The date of the latest certification of this metadata file in order to confirm that the metadata file produced is still up-to-date. Such a certification can also be done if the contents of the metadata file has not been amended.
2.2	Metadata last posted	META_POSTED	Date of the latest dissemination of the metadata.	Date	The date when this metadata file is disseminated (for Eurostat: by ESS Metadata Handler).
2.3	Metadata last update	META_LAST_UPDATE	Date of last update of the content of the metadata.	Date	The date when this metadata file is last updated (for Eurostat: by ESS Metadata Handler).
3	**Statistical presentation**	STAT_PRES			
3.1	Data description	DATA_DESCR	Main characteristics of the data set described in an easily understandable manner, referring to the data and indicators disseminated.	Text	Describe the main characteristics of the data set in an easily understandable manner, referring to the main data and indicators disseminated. This short description should be understood immediately and easily by the users.
3.2	Classification system	CLASS_SYSTEM	Arrangement or division of objects into groups based on characteristics which the objects have in common.	Text	List all classifications which are used for the data set produced (with their detailed names).
3.3	Sector coverage	COVERAGE_SECTOR	Main economic or other sectors covered by the statistics.	Text	List the main economic or other sectors covered by the data set produced, in also adding the size classes used.
3.4	Statistical concepts and definitions	STAT_CONC_DEF	Statistical characteristics of statistical observations.	Text	Describe in short the main statistical variables provided. The definition and types of variables provided should be listed, together with any information on discrepancies from the ESS/ international
3.5	Statistical unit	STAT_UNIT	Entity for which information is sought and for which statistics are ultimately compiled.	Text	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.
3.6	Statistical population	STAT_POP	The total membership or population or "universe" of a defined class of people, objects or events.	Text	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.
3.7	Reference area	REF_AREA	The country or geographic area to which the measured statistical phenomenon relates.	Text/Coded (code list: CL_REF_AREA)	**At European level:** The geographical area covered by the data set disseminated (e.g. EU Members states, EU regions, USA, Japan, etc. as well as aggregates such as EU-28, EEA). **At national level:** the country, the regions and aggregates covered by the data set disseminated.
3.8	Time coverage	COVERAGE_TIME	The length of time for which data are available.	Text	The time periods covered by the data set should be described (i.e. the length of time for which data set is disseminated, e.g. from 1985 to 2006 for certain annual data).

Figure 3: Fragment of ESMS specification

[17] Eurostat. Euro-SDMX Metadata Structure (ESMS). http://ec.europa.eu/eurostat/data/metadata

Another standardised metadata structure currently in use is the "ESS Standard Quality Report Structure" (ESQRS). Quality reports are produced and partly disseminated by Eurostat in this format. Eurostat has also recently introduced a "Single Integrated Metadata Structure" (SIMS), which represents the union of referential metadata attributes coming from ESMS and ESQRS, providing an integrated framework of concepts on quality assessment and more general reference metadata.

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). DCAT was developed in the period from June 2011 through December 2013 by the Government Linked Data Working Group[18]. 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.[19]

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*
- 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 4.

[18] W3C. Government Linked Data Working Group. https://www.w3.org/2011/gld/charter
[19] US spelling from the original.

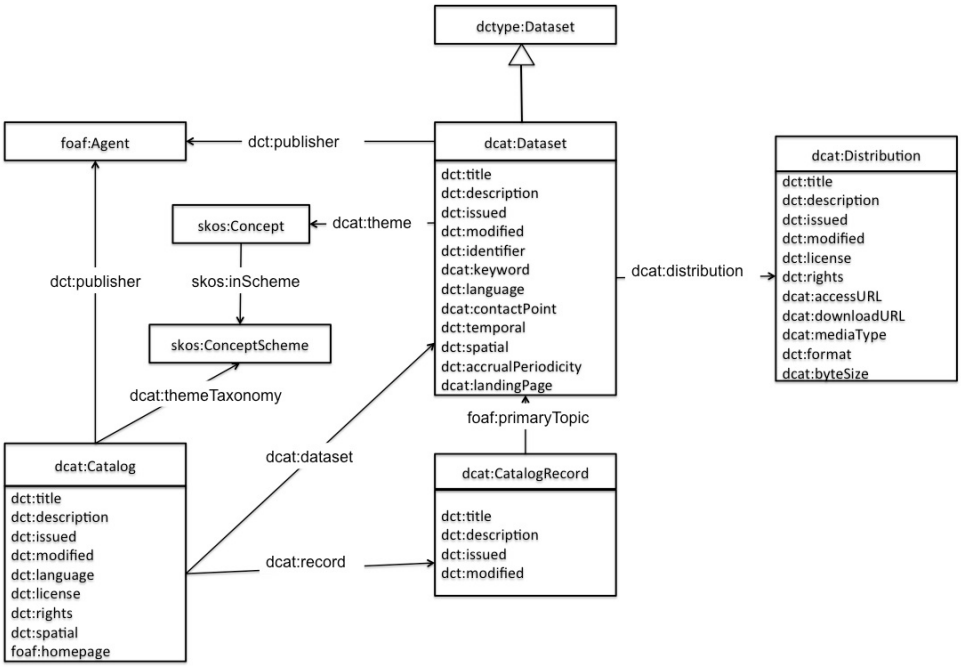


Figure 4: DCAT schematic data model

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)[20] of the Application Profile was published in September 2013. In 2015, a revised version (1.1)[21] 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 5.

[20] 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

[21] 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

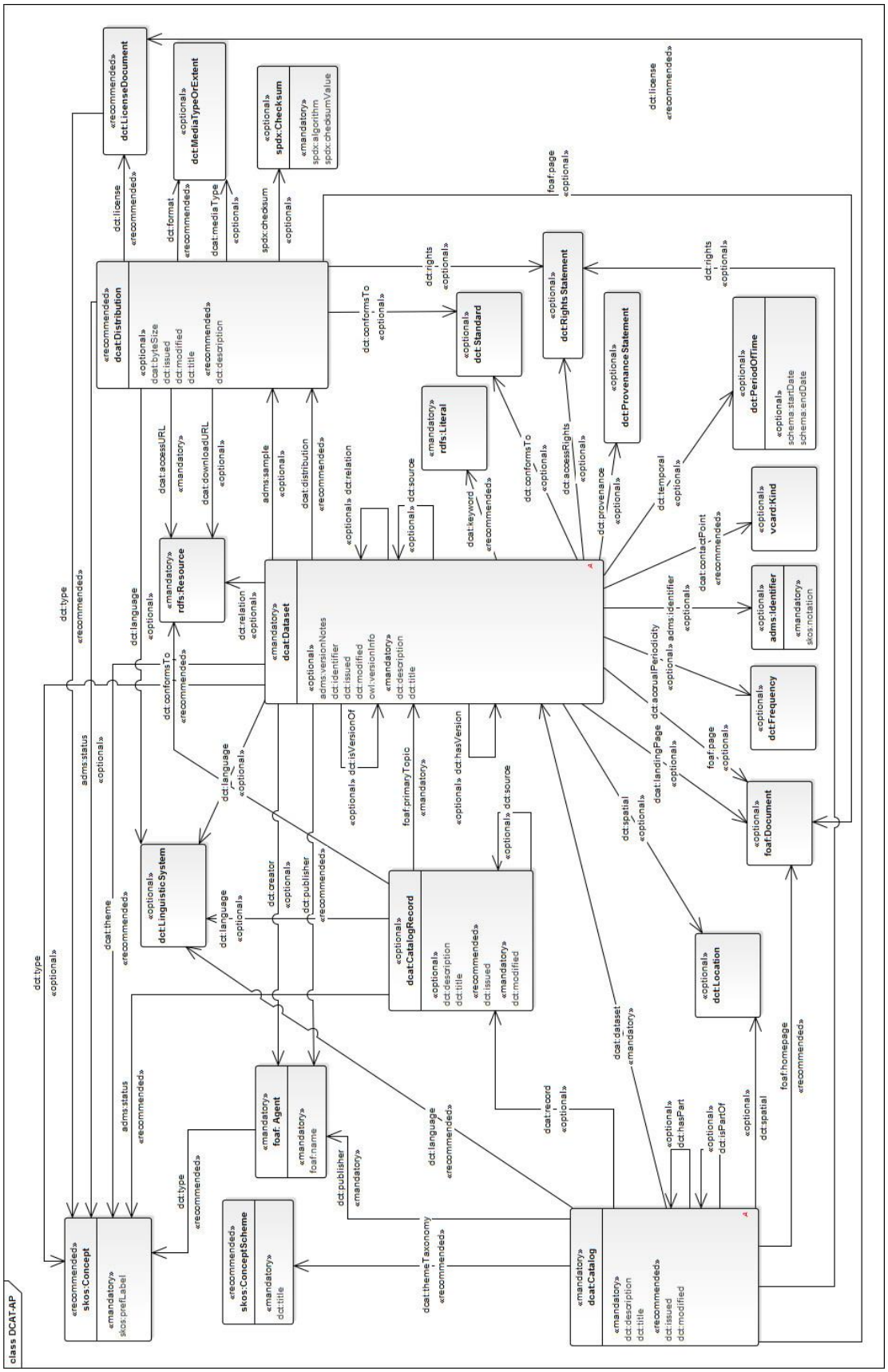


Figure 5: DCAT-AP Data Model

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[22]).

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)[23] of the GeoDCAT-AP was published in December 2015.

[22] http://publications.europa.eu/mdr/eurovoc/
[23] 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 is organised around the concept of the qb:DataSet, which is defined as *a collection of statistical data that corresponds to a defined structure*. The concept of a Dataset in DCAT (and DCAT-AP and StatDCAT-AP) is more generally defined as *a collection of data*. So the main distinction is that DCAT is concerned with the overall characteristics of a dataset, while the Data Cube Vocabulary is concerned with the structure of the data itself. The specification of the Data Cube Vocabulary mentions the use of Dublin Core and DCAT terms to describe the overall characteristics of Data Cube datasets.

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 builds upon the core of the SDMX Information Model with its concepts of dimensions, attributes and measures described in a data structure definition. 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 also recognises the use of SDMX cross-domain concepts and code lists on which several statistical data and metadata structures are being standardised.

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

An overview of its key terms and their relationships is shown in Figure 6.

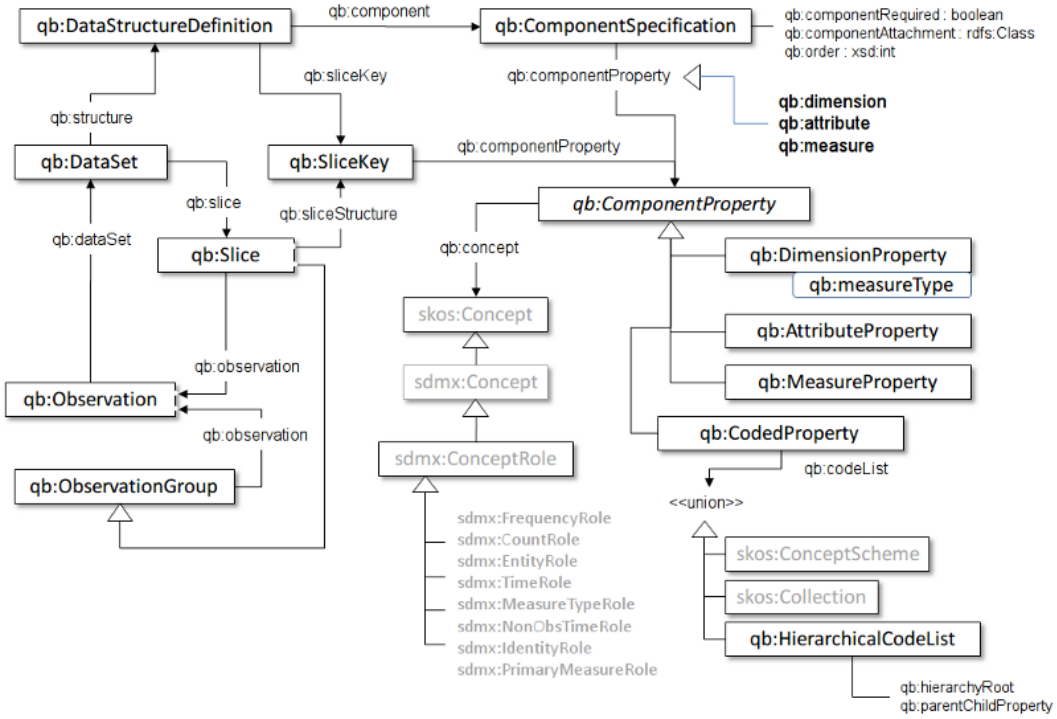


Figure 6: Data Cube vocabulary overview of key terms and relationships

4 USE CASES

4.1 Improve discoverability of statistical datasets on open data portals

Within the EU, Eurostat is the organization having the 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 profile 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 that 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 values. For that the W3C vocabulary Data Cube[24] exists. Work on StatDCAT-AP may, however, include discussions at this level since it may improve insight.

[24] 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[25] 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 principles underlying the work on StatDCAT-AP are:

- align with DCAT and DCAT-AP
- focus primarily on metadata elements that contribute to data discovery
- using metadata terms from existing, well-known and well-maintained vocabularies, including ISA Core Vocabularies and Eurostat metadata vocabularies
- encourage the use of common controlled vocabularies, preferably ones maintained in the Metadata Registry (MDR) by the 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.

[25] 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)[26].

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 particularly 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 5 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[27] 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.

[26] 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
[27] W3C. Data Catalogue Vocabulary. W3C Recommendation 16 January 2014. https://www.w3.org/TR/vocab-dcat/

6.2 Extensions and specific usage for description of statistical datasets

Discussions during the development of the StatDCAT-AP specifications brought out a number of requirements for the description of statistical datasets that were not met by existing properties in DCAT-AP. The following sections present the extensions that are included in StatDCAT-AP to meet those requirements. Some of the extensions are re-used from existing RDF vocabularies, others are defined in a new namespace specific for StatDCAT-AP. The namespace URI for this StatDCAT-AP namespace is http://data.europa.eu/(xyz)/statdcat-ap/. The string (xyz) will be assigned by the URI Committee responsible for the management of the persistent URIs of the EU institutions and bodies.

All issues discussed in the course of the development of StatDCAT-AP can be seen at https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/issue/all.

6.2.1 Dimensions an attributes

A requirement has been expressed to expose information about:

- Attributes: components used to qualify and interpret observed values such as units of measure, scaling factors
- Dimensions: components that identify observations such as time, sex, age, regions

The following two properties are created in the StatDCAT-AP namespace:

Property	Attribute
URI	stat:attribute
Range	qb:AttributeProperty, expressed as a URI.
Definition	A component used to qualify and interpret observed values
Comment	Attributes enable specification of the units of measure, any scaling factors and metadata such as the status of the observation (e.g. estimated, provisional).
Usage note	This property is to be used to identify an attribute that is used in the Dataset. The attribute is a shortcut for the property path qb:structure -> qb:component -> qb:attribute. If a Dataset is modelled using the Data Cube vocabulary, the value of the dq:attribute property should be used as the value for this property.

Property	Dimension
URI	stat:dimension
Range	qb:DimensionProperty, expressed as a URI.
Definition	A component that identifies observations
Comment	Examples of dimensions include the time to which the observation applies, or a geographic region which the observation covers.
Usage note	This property is to be used to identify a dimension that is used in the Dataset. The attribute is a shortcut for the property path qb:structure -> qb:component -> qb:dimension.

Property	Dimension
	If a Dataset is modelled using the Data Cube vocabulary, the value of the dq:dimension property should be used as the value for this property.

Example:

A dataset, provided by a national statistical portal, contains employment data concerning all five regions of the country, based on sex and age.

In the sample below, the property stat:dimension is used to represent the StatDCAT-AP descriptors of the observations contained in the dataset, in this case age, sex and region. Similarly, the property stat:attribute is used to represent the unit of measure, in this case percentage of employment based on the mentioned characteristics.

<http://nationalportal.org/data/employment> a dcat:Dataset

stat:dimension <http://nationalportal.org/dimension/age#>;

stat:dimension <http://nationalportal.org/dimension/sex#>;

stat:dimension <http://nationalportal.org/dimension/geo#>;

stat:attribute <http://nationalportal.org/attribute/perc#>.

6.2.2 Quality aspects

Quality aspects are very important for datasets in general and for statistical datasets in particular. The current specification includes a first approach that allows certain annotations related to quality to be made, either through linking to quality information that is already published elsewhere or by including text with quality information.

The following annotation property is included, re-used from the Data Quality Vocabulary[28] that is being developed by the Data on the Web Working Group at W3C.

Property	Quality annotation
URI	dqv:hasQualityAnnotation
Range	oa:Annotation
Definition	A statement related to quality of the Dataset, including rating, quality certificate, feedback that can be associated to datasets or distributions.
Comment	The information may include quality aspects such as accuracy, reliability, comparability, coherence, relevance, timeliness etc.
Usage note	The annotation requires the provision of information about the motivation of the annotation (oa:motivation), and an explicit link to the resource being annotated (oa:hasTarget) together with either a link to a resource that contains the annotation (oa:hasBody) or text filed (oa:bodyText).

Example:

In the sample code below is described the quality of information contained using the dqv:hasQualityAnnotation property of StatDCAT-AP.

<http://qualifications.org/QualityCertificate1> a oa:Annotation.

<http://nationalportal.org/data/employment> a dcat:Dataset;

dqv:hasQualityAnnotation <http://qualifications.org/QualityCertificate1>.

<http://qualifications.org/QualityCertificate1> a dqv:QualityCertificate;

oa:hasBody <http://qualifications.org/QualityCertificate1/body>.

[28] Data on the Web Best Practices: Data Quality Vocabulary. W3C Working Draft 19 May 2016. https://www.w3.org/TR/vocab-dqv/

Another property is created in the StatDCAT-AP namespace.

Property	Unit of measurement
URI	http://data.europa.eu/(xyz)/statdcat-ap/statMeasure
Range	skos:Concept
Definition	A unit of measurement of the observations in the dataset
Comment	Examples are Euro, square kilometre, purchasing power standard (PPS), full-time equivalent, percentage. Values should be taken from a controlled vocabulary, possible provided as an MDR authority.

Example:

In the sample code below the unit of measurement, in this case 'percentage', is described using the property stat:statMeasure of StatDCAT-AP.

<http://nationalportal.org/data/employment> a dcat:Dataset;

stat:statMeasure <http://example.com/measures/percentage> .

In future work, a more fundamental treatment of quality aspects may be undertaken, for example on the basis of the work done at Eurostat on the Single Integrated Metadata Structure.

6.2.3 Visualisation

One of the requirements raised during discussions was the need to be able to link to a visualisation of the data, for example a document or Webpage where a tabular or graphical representation of the data can be viewed, or an interactive service where the data can be accessed and viewed.

The agreed approach for these types of visualisations is to model them as Distributions with a type of "visualisation" from the MDR Distribution type Named Authority List[29].

In order to implement this approach, a property to indicate the type of distribution is added to the class Distribution, re-used from Dublin Core.

Property	Type of distribution
URI	http://purl.org/dc/type
Range	URI of a term in a controlled vocabulary
Definition	The nature or genre of the resource.
Comment	Recommended best practice is to use a controlled vocabulary
Usage note	This property is to be used to indicate the type of a Distribution, in particular when the Distribution is a visualisation. For visualisations, use the term 'Visualisation" From the MDR Distribution type Named Authority List

Example:

In the sample code below the type of a distribution, in this case a visualisation of the dataset, is described using the dct:type property of StatDCAT-AP.

<http://nationalportal.org/data/employment/emplvis> a dcat:Distribution.

[29] MDR Authorities. Distribution types. http://publications.europa.eu/mdr/authority/distribution-type/

```
dct:type <http://publications.europa.eu/resource/authority/documentation-type/VISUALIZATION> ;
dcat:accessURL <http://nationalportal.org/data/employment/emplvis.html> .
```

Further use of the Type property on Distribution may be considered in the future, for example to indicate that data can be accessed through a service.

6.2.4 Other extensions

A requirement was brought forward to allow expression of the number of data series contained in a dataset.

A series is a unique cross product of values of dimensions (excluding time). The number of data series therefore gives an indication of the potential scope of a data set. 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 created in the StatDCAT-AP namespace.

Property	Number of data series
URI	http://data.europa.eu/(xyz)/statdcat-ap/numSeries
Subproperty of	dct:extent
Range	dct:SizeOrDuration, expressed as xsd:integer
Definition	The number of data series in a Dataset
Comment	The number of data series provides information on how values in the Dataset are grouped

Example:

The sample code below describes the number of data series contained in the dataset using the stat:numSeries property of StatDCAT-AP.

```
<http://nationalportal.org/data/employment> a dcat:Dataset;
    stat:numSeries '5'^^xsd:integer.
```

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[30].

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.).

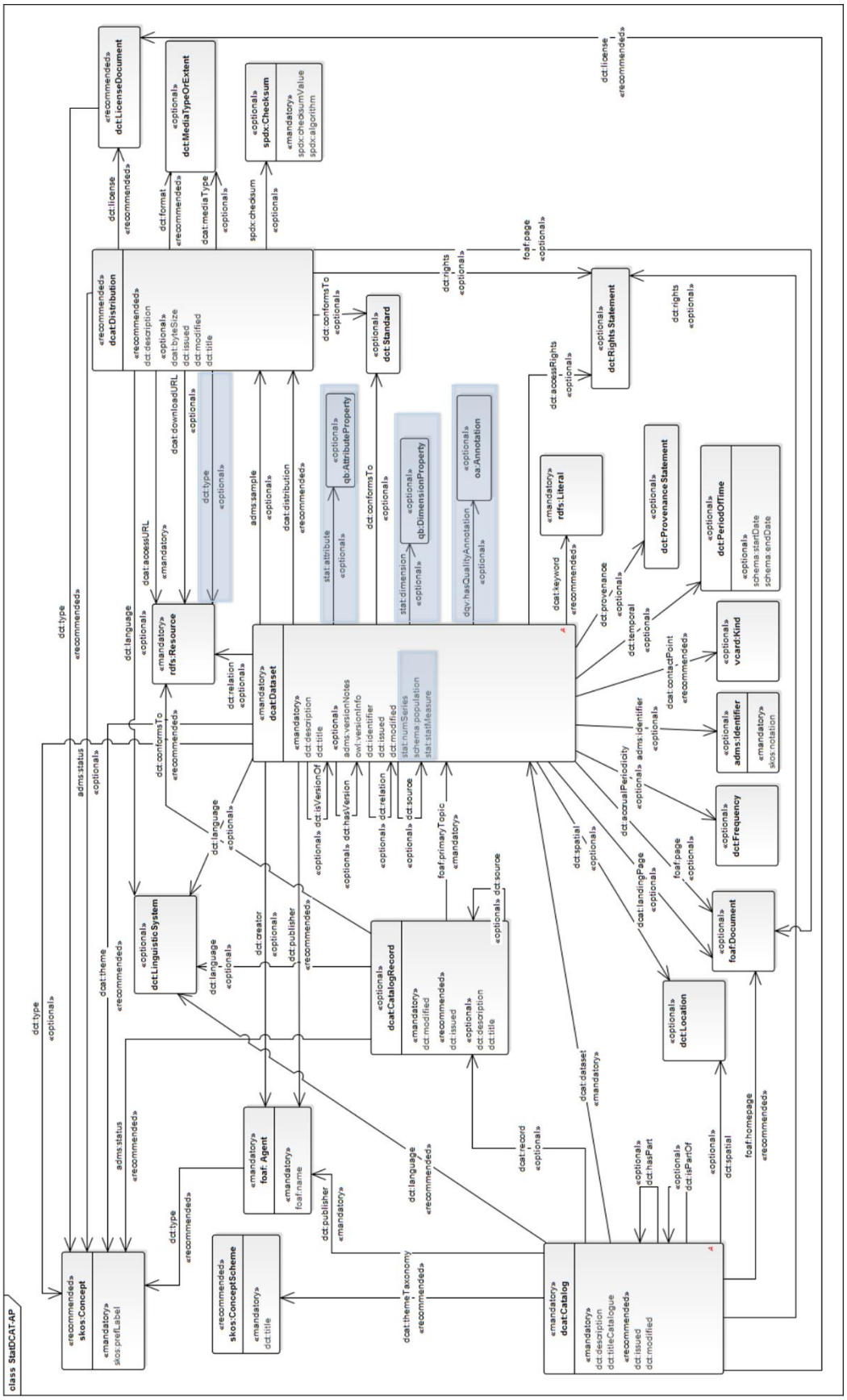
6.4 Namespaces

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#	**dct**: http://purl.org/dc/terms/
dcat: http://www.w3.org/ns/dcat#	**skos**: http://www.w3.org/2004/02/skos/core#
foaf: http://xmlns.com/foaf/0.1/	**spdx**: http://spdx.org/rdf/terms#
owl: http://www.w3.org/2002/07/owl#	stat: http://data.europa.eu/(xyz)/statdcat-ap/
qb: http://purl.org/linked-data/cube#	**xsd**: http://www.w3.org/2001/XMLSchema#
rdfs: http://www.w3.org/2000/01/rdf-schema#	**vcard**: http://www.w3.org/2006/vcard/ns#
schema: http://schema.org/	

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

6.5 UML Class diagram



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[31] 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[32].	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
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, including visualisations of the data	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

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

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[33] 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
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
Provenance Statement	A statement of any changes in ownership and custody of a resource since its creation that are significant for its authenticity, integrity, and interpretation	dct:ProvenanceStatement	http://dublincore.org/documents/dcmi-terms/#terms-ProvenanceStatement

Additional optional classes for StatDCAT-AP

Class name	Usage note for the Application Profile	URI	Reference
Annotation	A statement that annotates a resource. In this profile, used for statements related to quality of the Dataset, including rating, quality certificate, feedback that can be associated to datasets or distributions.	oa:Annotation	https://www.w3.org/TR/2016/WD-annotation-vocab-20160331/#annotation
Attribute Property	A component property which represent an attribute of observations in the Dataset, e.g. unit of measurement	qb:AttributeProperty	https://www.w3.org/TR/vocab-data-cube/#ref_qb_AttributeProperty

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

Dimension Property	A component property which represent a dimension in the Dataset	qb:DimensionProperty	https://www.w3.org/TR/vocab-data-cube/#ref_qb_DimensionProperty
Size or duration	A dimension or extent, e.g. the number of data series in a Dataset	dct:SizeOrDuration	http://dublincore.org/documents/dcmi-terms/#terms-SizeOrDuration

6.7 Description of properties per class

The following section describes the properties to be used in StatDCAT-AP. It contains the specification for DCAT-AP and indicates the extensions for StatDCAT-AP separately under the relevant classes. The extensions are described in section 6.2.

6.7.1 Catalogue

StatDCAT-AP does not specify additional properties for Catalogue, comparing to the ones used in DCAT-AP 1.1

Mandatory properties for Catalogue

Property	URI	Range	Usage note	Card.
dataset	dcat:dataset	dcat:Dataset	This property links the Catalogue with a Dataset that is part of the Catalogue.	1..n
description	dct:description	rdfs:Literal	This property contains a free-text account of the Catalogue. This property can be repeated for parallel language versions of the description. For further information on multilingual issues, please refer to section 10.	1..n
publisher	dct:publisher	foaf:Agent	This property refers to an entity (organisation) responsible for making the Catalogue available.	1..1
title	dct:title	rdfs:Literal	This property contains a name given to the Catalogue. This property can be repeated for parallel language versions of the name.	1..n

Recommended properties for Catalogue

Property	URI	Range	Usage note	Card.
homepage	foaf:homepage	foaf:Document	This property refers to a web page that acts as the main page for the Catalogue.	0..1
language	dct:language	dct:LinguisticSystem	This property refers to a language used in the textual metadata describing titles, descriptions, etc. of the Datasets in the Catalogue. This property can be repeated if the metadata is provided in multiple languages.	0..n
licence	dct:license	dct:LicenseDocument	This property refers to the licence under which the Catalogue can be used or reused.	0..1
release date	dct:issued	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the date of formal issuance (e.g., publication) of the Catalogue.	0..1
themes	dcat:themeTaxonomy	skos:ConceptScheme	This property refers to a knowledge organization system used to classify the Catalogue's Datasets.	0..n
update/ modification date	dct:modified	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the most recent date on which the Catalogue was modified.	0..1

Optional properties for Catalogue

Property	URI	Range	Usage note	Card.
has part	dct:hasPart	dcat:Catalog	This property refers to a related Catalogue that is part of the described Catalogue	0..n
is part of	dct:isPartOf	dcat:Catalog	This property refers to a related Catalogue in which the described Catalogue is physically or logically included.	0..1
record	dcat:record	dcat:CatalogRecord	This property refers to a Catalogue Record that is part of the Catalogue	0..n
rights	dct:rights	dct:RightsStatement	This property refers to a statement that specifies rights associated with the Catalogue.	0..1

spatial / geographic	dct:spatial	dct:Location	This property refers to a geographical area covered by the Catalogue.	0..n

6.7.2 Catalogue Record

StatDCAT-AP does not specify additional properties for Catalogue Record, comparing to the ones used in DCAT-AP 1.1

Mandatory properties for Catalogue Record

Property	URI	Range	Usage note	Card.
primary topic	foaf:primaryTopic	dcat:Dataset	This property links the Catalogue Record to the Dataset described in the record.	1..1
update/ modification date	dct:modified	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the most recent date on which the Catalogue entry was changed or modified.	1..1

Recommended properties for Catalogue Record

Property	URI	Range	Usage note	Card.
application profile	dct:conformsTo	rdfs:Resource	This property refers to an Application Profile that the Dataset's metadata conforms to	0..1
change type	adms:status	skos:Concept	This property refers to the type of the *latest* revision of a Dataset's entry in the Catalogue. It MUST take one of the values :created, :updated or :deleted depending on whether this *latest* revision is a result of a creation, update or deletion.	0..1
listing date	dct:issued	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the date on which the description of the Dataset was included in the Catalogue.	0..1

Optional properties for Catalogue Record

Property	URI	Range	Usage note	Card.
description	dct:description	rdfs:Literal	This property contains a free-text account of the record. This property can be repeated for parallel language versions of the description.	0..n
language	dct:language	dct:LinguisticSystem	This property refers to a language used in the textual metadata describing titles, descriptions, etc. of the Dataset. This property can be repeated if the metadata is provided in multiple languages.	0..n
source metadata	dct:source	dcat:CatalogRecord	This property refers to the original metadata that was used in creating metadata for the Dataset	0..1
title	dct:title	rdfs:Literal	This property contains a name given to the Catalogue Record. This property can be repeated for parallel language versions of the name.	0..n

6.7.3 Dataset

On top of the properties used in DCAT-AP 1.1, StatDCAT-AP specifies six additional properties for Dataset.

Mandatory properties for Dataset

Property	URI	Range	Usage note	Card
description	dct:description	rdfs:Literal	This property contains a free-text account of the Dataset. This property can be repeated for parallel language versions of the description.	1..n
title	dct:title	rdfs:Literal	This property contains a name given to the Dataset. This property can be repeated for parallel language versions of the name.	1..n

Recommended properties for Dataset

Property	URI	Range	Usage note	Card
contact point	dcat:contactPoint	vcard:Kind	This property contains contact information that can be used for sending comments about the Dataset.	0..n
dataset distribution	dcat:distribution	dcat:Distribution	This property links the Dataset to an available Distribution.	0..n
keyword/ tag	dcat:keyword	rdfs:Literal	This property contains a keyword or tag describing the Dataset.	0..n
publisher	dct:publisher	foaf:Agent	This property refers to an entity (organisation) responsible for making the Dataset available.	0..1
theme/ category	dcat:theme, subproperty of dct:subject	skos:Concept	This property refers to a category of the Dataset. A Dataset may be associated with multiple themes.	0..n

Optional properties for Dataset

Property	URI	Range	Usage note	Card.
access rights	dct:accessRights	dct:RightsStatement	This property refers to information that indicates whether the Dataset is open data, has access restrictions or is not public. A controlled vocabulary with three members (:public, :restricted, :non-public) will be created and maintained by the Publications Office of the EU.	0..1
conforms to	dct:conformsTo	dct:Standard	This property refers to an implementing rule or other specification.	0..n
documentation	foaf:page	foaf:Document	This property refers to a page or document about this Dataset.	0..n
frequency	dct:accrualPeriodicity	dct:Frequency	This property refers to the frequency at which the Dataset is updated.	0..1
has version	dct:hasVersion	dcat:Dataset	This property refers to a related Dataset that is a version, edition, or adaptation of the described Dataset.	0..n
identifier	dct:identifier	rdfs:Literal	This property contains the main identifier for the Dataset, e.g. the URI or other unique identifier in the context of the Catalogue.	0..n
is version of	dct:isVersionOf	dcat:Dataset	This property refers to a related Dataset of which the described Dataset is a version, edition, or adaptation.	0..n
landing page	dcat:landingPage	foaf:Document	This property refers to a web page that provides access to the Dataset, its Distributions and/or additional information. It is intended to point to a landing page at the original data provider, not to a page on a site of a third party, such as an aggregator.	0..n
language	dct:language	dct:LinguisticSystem	This property refers to a language of the Dataset. This property can be repeated if there are multiple languages in the Dataset.	0..n
other identifier	adms:identifier	adms:Identifier	This property refers to a secondary identifier of the Dataset, such as MAST/ADS[34], DataCite[35], DOI[36], EZID[37] or W3ID[38].	0..n
provenance	dct:provenance	dct:ProvenanceStatement	This property contains a statement about the lineage of a Dataset.	0..n
related resource	dct:relation	rdfs:Resource	This property refers to a related resource.	0..n
release date	dct:issued	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the date of formal issuance (e.g., publication) of the Dataset.	0..1

[34] Mikulski Archive for Space Telescopes (MAST). Referencing Data Sets in Astronomical Literature. http://archive.stsci.edu/pub_dsn.html
[35] DataCite. http://www.datacite.org/
[36] DOI. Digital Object Identifier. http://www.doi.org/
[37] EZID. http://n2t.net/ezid
[38] W3C Permanent Identifier Community Group. Permanent Identifiers for the Web. https://w3id.org/

Property	URI	Range	Usage note	Card.
sample	adms:sample	dcat:Distribution	This property refers to a sample distribution of the dataset	0..n
source	dct:source	dcat:Dataset	This property refers to a related Dataset from which the described Dataset is derived.	0..n
spatial/ geographical coverage	dct:spatial	dct:Location	This property refers to a geographic region that is covered by the Dataset.	0..n
temporal coverage	dct:temporal	dct:PeriodOfTime	This property refers to a temporal period that the Dataset covers.	0..n
type	dct:type	skos:Concept	This property refers to the type of the Dataset. A controlled vocabulary for the values has not been established.	0..1
update/ modification date	dct:modified	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the most recent date on which the Dataset was changed or modified.	0..1
version	owl:versionInfo	rdfs:Literal	This property contains a version number or other version designation of the Dataset.	0..1
version notes	adms:versionNotes	rdfs:Literal	This property contains a description of the differences between this version and a previous version of the Dataset. This property can be repeated for parallel language versions of the version notes.	0..n

Additional optional properties for Dataset

Property	URI	Range	Usage note	Card
attribute	stat:attribute	qb:AttributeProperty	This property links to a component used to qualify and interpret observed values, e.g. units of measure, any scaling factors and metadata such as the status of the observation (e.g. estimated, provisional).	0..n
dimension	stat:dimension	qb:DimensionProperty	This property links to a component that identifies observations, e.g. the time to which the observation applies, or a geographic region which the observation covers	0..n
number of data series	stat:numSeries	rdfs:Literal typed as xsd:integer	This property contains the number of data series contained in the Dataset	
quality annotation	dqv:hasQualityAnnotation	oa:Annotation	This property links to a statement related to quality of the Dataset, including rating, quality certificate, feedback that can be associated to the Dataset.	0..n
unit of measurement	stat:statMeasure	skos:Concept	This property links to a unit of measurement of the observations in the dataset, for example Euro, square kilometre, purchasing power standard (PPS), full-time equivalent, percentage.	0..n

6.7.4 Distribution

On top of the properties used in DCAT-AP 1.1, StatDCAT-AP specifies one additional property for Distribution.

Mandatory properties for Distribution

Property	URI	Range	Usage note	Card
access URL	dcat:accessURL	rdfs:Resource	This property contains a URL that gives access to a Distribution of the Dataset. The resource at the access URL may contain information about how to get the Dataset.	1..n

Recommended properties for Distribution

Property	URI	Range	Usage note	Card
description	dct:description	rdfs:Literal	This property contains a free-text account of the Distribution. This property can be repeated for parallel language versions of the description.	0..n
format	dct:format	dct:MediaTypeOrExtent	This property refers to the file format of the Distribution.	0..1
licence	dct:license	dct:LicenseDocument	This property refers to the licence under which the Distribution is made available.	0..1

Optional properties for Distribution

Property	URI	Range	Usage note	Card.
byte size	dcat:byteSize	rdfs:Literal typed as xsd:decimal	This property contains the size of a Distribution in bytes.	0..1
checksum	spdx:checksum	spdx:Checksum	This property provides a mechanism that can be used to verify that the contents of a distribution have not changed	0..1
documentation	foaf:page	foaf:Document	This property refers to a page or document about this Distribution.	0..n
download URL	dcat:downloadURL	rdfs:Resource	This property contains a URL that is a direct link to a downloadable file in a given format.	0..n
language	dct:language	dct:LinguisticSystem	This property refers to a language used in the Distribution. This property can be repeated if the metadata is provided in multiple languages.	0..n
linked schemas	dct:conformsTo	dct:Standard	This property refers to an established schema to which the described Distribution conforms.	0..n
media type	dcat:mediaType, subproperty of dct:format	dct:MediaTypeOrExtent	This property refers to the media type of the Distribution as defined in the official register of media types managed by IANA.	0..1
release date	dct:issued	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the date of formal issuance (e.g., publication) of the Distribution.	0..1
rights	dct:rights	dct:RightsStatement	This property refers to a statement that specifies rights associated with the Distribution.	0..1
status	adms:status	skos:Concept	This property refers to the maturity of the Distribution	0..1
title	dct:title	rdfs:Literal	This property contains a name given to the Distribution. This property can be repeated for parallel language versions of the description.	0..n
update/ modification date	dct:modified	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the most recent date on which the Distribution was changed or modified.	0..1

Additional optional property for Distribution

Property	URI	Range	Usage note	Card
type	dct:type	rdfs:Resource	This property links to a type of the Distribution, e.g. that it is a visualisation	0..1

6.7.5 Agent

Mandatory properties for Agent

Property	URI	Range	Usage note	Card.
name	foaf:name	rdfs:Literal	This property contains a name of the agent. This property can be repeated for different versions of the name (e.g. the name in different languages)	1..n

Recommended properties for Agent

Property	URI	Range	Usage note	Card.
type	dct:type	skos:Concept	This property refers to a type of the agent that makes the Catalogue or Dataset available	0..1

6.7.6 Category Scheme

Mandatory properties for Category Scheme

Property	URI	Range	Usage note	Card.
title	dct:title	rdfs:Literal	This property contains a name of the category scheme. May be repeated for different versions of the name	1..n

6.7.7 Category

Mandatory properties for Category

Property	URI	Range	Usage note	Card.
preferred label	skos:prefLabel	rdfs:Literal	This property contains a preferred label of the category. This property can be repeated for parallel language versions of the label.	1..n

6.7.8 Checksum

Mandatory properties for Checksum

Property	URI	Range	Usage note	Card.
algorithm	spdx:algorithm	spdx:checksumAlgorithm_sha1	This property identifies the algorithm used to produce the subject Checksum. Currently, SHA-1 is the only supported algorithm. It is anticipated that other algorithms will be supported at a later time.	1..1
checksum value	spdx:checksumValue	rdfs:Literal typed as xsd:hexBinary	This property provides a lower case hexadecimal encoded digest value produced using a specific algorithm.	1..1

6.7.9 Identifier

Mandatory properties for Identifier

Property	URI	Range	Usage note	Card.
notation	skos:notation	rdfs:Literal typed with the URI of one of the members of the DataCite Resource Identifier Scheme[39]	This property contains a string that is an identifier in the context of the identifier scheme referenced by its datatype.	0..1

6.7.10 Licence Document

Recommended properties for Licence Document

Property	URI	Range	Usage note	Card.
licence type	dct:type	skos:Concept	This property refers to a type of licence, e.g. indicating 'public domain' or 'royalties required'.	0..1

[39] DataCite Resource Identifier Scheme. http://purl.org/spar/datacite/ResourceIdentifierScheme

6.7.11 Period of Time

Optional properties for Period of Time

Property	URI	Range	Usage note	Card.
start date/time	schema:startDate	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the start of the period	0..1
end date/time	schema:endDate	rdfs:Literal typed as xsd:date or xsd:dateTime	This property contains the end of the period	0..1
Please note that while both properties are optional, one of the two must be present for each instance of the class dct:PeriodOfTime, if such an instance is present. The start of the period should be understood as the start of the date, hour, minute etc. given (e.g. starting at midnight at the beginning of the day if the value is a date); the end of the period should be understood as the end of the date, hour, minute etc. given (e.g. ending at midnight at the end of the day if the value is a date)				

6.8 Controlled vocabularies

StatDCAT-AP uses the same controlled vocabularies as DCAT-AP. Section 6.8.2 specifies the controlled vocabularies to be used, while section 6.8.3 provides a table with the mappings from the Eurostat theme vocabulary to the MDR data themes vocabulary.

6.8.1 Requirements for controlled vocabularies

The following is a list of requirements that were identified for the controlled vocabularies to be recommended in this Application Profile.

Controlled vocabularies SHOULD:

- Be published under an open licence.
- Be operated and/or maintained by an institution of the European Union, by a recognised standards organisation or another trusted organisation.
- Be properly documented.
- Have labels in multiple languages, ideally in all official languages of the European Union.
- Contain a relatively small number of terms (e.g. 10-25) that are general enough to enable a wide range of resources to be classified.
- Have terms that are identified by URIs with each URI resolving to documentation about the term.
- Have associated persistence and versioning policies.

These criteria do not intend to define a set of requirements for controlled vocabularies in general; they are only intended to be used for the selection of the controlled vocabularies that are proposed for this Application Profile.

6.8.2 Controlled vocabularies to be used

In the table below, a number of properties are listed with controlled vocabularies that MUST be used for the listed properties. The declaration of the following controlled vocabularies as mandatory ensures a minimum level of interoperability.

Property URI	Used for Class	Vocabulary name	Vocabulary URI	Usage note
dcat:mediaType	Distribution	IANA Media Types[40]	http://www.iana.org/assignments/media-types/media-types.xhtml	
dcat:theme	Dataset	MDR Dataset Theme Vocabulary	http://publications.europa.eu/resource/authority/data-theme and http://publications.europa.eu/mdr/authority/data-theme/	The values to be used for this property are the URIs of the concepts in the vocabulary.
dcat:themeTaxonomy	Catalogue	MDR Dataset Theme Vocabulary	http://publications.europa.eu/resource/authority/data-theme and http://publications.europa.eu/mdr/authority/data-theme/	The value to be used for this property is the URI of the vocabulary itself, i.e. the concept scheme, not the URIs of the concepts in the vocabulary.
dct:accrualPeriodicity	Dataset	MDR Frequency Named Authority List[41]	http://publications.europa.eu/mdr/authority/frequency	
dct:format	Distribution	MDR File Type Named Authority List[42]	http://publications.europa.eu/mdr/authority/file-type/	
dct:language	Catalogue, Dataset	MDR Languages Named Authority List[43]	http://publications.europa.eu/mdr/authority/language/	
dct:publisher	Catalogue, Dataset	MDR Corporate Bodies Named Authority List[44]	http://publications.europa.eu/mdr/authority/corporate-body/	The Corporate bodies NAL must be used for European institutions and a small set of international organisations. In case of other types of organisations, national, regional or local vocabularies should be used.
dct:spatial	Catalogue, Dataset	MDR Continents Named Authority List[45], MDR Countries Named Authority List[46], MDR Places Named Authority List[47], Geonames	http://publications.europa.eu/mdr/authority/country/, http://publications.europa.eu/mdr/authority/place/, http://publications.europa.eu/mdr/authority/continent/, http://sws.geonames.org/	The MDR Name Authority Lists must be used for continents, countries and places that are in those lists; if a particular location is not in one of the mentioned Named Authority Lists, Geonames URIs must be used.
adms:status	Catalogue Record	ADMS change type vocabulary	http://purl.org/adms/changetype/	:created, :updated, :deleted
adms:status	Distribution	ADMS status vocabulary	http://purl.org/adms/status/	The list of terms in the ADMS status vocabulary is included in the ADMS specification[48]
dct:type	Agent	ADMS publisher type vocabulary	http://purl.org/adms/publishertype/	The list of terms in the ADMS publisher type vocabulary is included in the ADMS specification
dct:type	Licence Document	ADMS licence type vocabulary	http://purl.org/adms/licencetype/	The list of terms in the ADMS licence type vocabulary is

[40] Internet Assigned Numbers Authority (IANA). Media Types http://www.iana.org/assignments/media-types/
[41] Publications Office of the European Union. Metadata Registry. Authorities. Frequencies. http://publications.europa.eu/mdr/authority/frequency
[42] Publications Office of the European Union. Metadata Registry. Authorities. File types. http://publications.europa.eu/mdr/authority/file-type/
[43] Publications Office of the European Union. Metadata Registry. Authorities. Languages. http://publications.europa.eu/mdr/authority/language/
[44] Publications Office of the European Union. Metadata Registry. Authorities. Corporate bodies. http://publications.europa.eu/mdr/authority/corporate-body/
[45] Publications Office of the European Union. Metadata Registry. Authorities. Continents. http://publications.europa.eu/mdr/authority/continent/
[46] Publications Office of the European Union. Metadata Registry. Authorities. Countries. http://publications.europa.eu/mdr/authority/country/
[47] Publications Office of the European Union. Metadata Registry. Authorities. Places. http://publications.europa.eu/mdr/authority/place/
[48] European Commission. Joinup. Asset Description Metadata Schema (ADMS). ADMS 1.00. http://joinup.ec.europa.eu/asset/adms/release/100

				included in the ADMS specification

6.8.3 Mapping Eurostat theme vocabulary to the MDR data themes vocabulary

Eurostat theme	MDR data theme
General and regional statistics	No mapping
Economy and finance	Economy and finance
Population and social conditions	Population and society
Industry, trade and services	Economy and finance
Agriculture, forestry and fisheries	Agriculture, fisheries, forestry, foods
International trade	Economy and finance
Transport	Transport
Environment and energy	Environment / Energy
Science and technology	Science and technology

Notes:

- There is no mapping for the Eurostat theme 'General and regional statistics'
- Three Eurostat themes 'Economy and finance', 'Industry, trade and services' and 'International trade' all map to the single MDR data theme 'Economy and finance'
- The single Eurostat theme 'Environment and energy' maps to two MDR data themes 'Environment' and 'Energy'

6.8.4 Other controlled vocabularies

In addition to the proposed common vocabularies in section 6.8.2, which are mandatory to ensure minimal interoperability, implementers are encouraged to publish and to use further region or domain-specific vocabularies that are available online. While those may not be recognised by general implementations of the Application Profile, they may serve to increase interoperability across applications in the same region or domain. Examples are the full set of concepts in EuroVoc[49], the CERIF standard vocabularies[50], the Dewey Decimal Classification[51] and numerous other schemes.

6.8.5 Licence vocabularies

Concerning licence vocabularies, implementers are encouraged to use widely recognised licences such as Creative Commons licences[52], and in particular the CC Zero Public Domain Dedication[53], the Open Data Commons Public Domain Dedication and License

[49] EuroVoc. http://eurovoc.europa.eu/.
[50] http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.5/CERIF1.5_Semantics.xhtml
[51] OCLC. Dewey Summaries as Linked Data. http://www.oclc.org/dewey/webservices.en.html and http://dewey.info/
[52] Creative Commons. About The Licenses. http://creativecommons.org/licenses/
[53] Creative Commons. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. http://creativecommons.org/publicdomain/zero/1.0/

(PDDL)[54], the ISA Open Metadata Licence[55], the European Union Public Licence (EUPL)[56] or an open government licence such as the UK Open Government Licence[57].

Further activities in this area are undertaken by the Open Data Institute[58] with the Open Data Rights Statement Vocabulary[59] and by the Open Digital Rights Language (ODRL) Initiative[60].

[54] Open Data Commons Public Domain Dedication and License (PDDL). http://opendatacommons.org/licenses/pddl/
[55] ISA Open Metadata Licence v1.1, https://joinup.ec.europa.eu/category/licence/isa-open-metadata-licence-v11
[56] European Commission. Joinup. Open Source Software. European Union Public Licence (EUPL). http://joinup.ec.europa.eu/software/page/eupl
[57] The National Archives. Open Government Licence for public sector information. http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/
[58] Open Data Institute. http://www.theodi.org/
[59] Open Data Institute. Open Data Rights Statement Vocabulary. http://schema.theodi.org/odrs/
[60] Open Digital Rights Language (ODRL) Initiative. http://www.w3.org/community/odrl/

7 MAPPING AND EXTRACTION APPROACHES

It is not expected that systems will implement StatDCAT-AP as a native format, at least not in the short term. As the StatDCAT-AP format is intended as a common target format for export of metadata that may exist in a variety of standard and local formats, 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 restrict in any way the approach that local implementers may want to use 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 who manage local systems that are based on SDMX (e.g. Eurostat and other statistical agencies) to map their metadata to a SDMX-based intermediary format.

Such a format may enable common approaches among SDMX implementers and may lower the threshold for the export of metadata 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 Annex II, Annex III and Annex IV.

8 CONFORMANCE STATEMENT

8.1 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 Controlled vocabularies in DCAT-AP, 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.

8.2 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.).

9 AGENT ROLES

The DCAT Application Profile specified in this document has a single property to relate an Agent (typically, an organisation) to a Dataset. The only such 'agent role' that can be expressed in the current version of the profile is through the property dct:publisher (http://purl.org/dc/terms/publisher), defined as "An entity responsible for making the dataset available". A second property is available in the DCAT recommendation, dcat:contactPoint (http://www.w3.org/TR/vocab-dcat/#Property:dataset_contactPoint), defined as "Link a dataset to relevant contact information which is provided using VCard", but this is not an agent role as the value of this property is contact data, rather than a representation of the organisation as such.

In specific cases, for example in exchanging data among domain-specific portals, it may be useful to express other, more specific agent roles. In such cases, extensions to the base profile may be defined using additional properties with more specific meanings.

Two possible approaches have been discussed, particular in the context of the development of the domain-specific GeoDCAT Application Profile, an extension of the base DCAT Application Profile.

The first possible approach is based on the use of a predicate vocabulary that provides a set of properties that represent additional types of relationships between Datasets and Agents. For example, properties could be defined, such as foo:owner, foo:curator or foo:responsibleParty, in addition to the use of existing well-known properties, such as dct:creator and dct:rightsHolder. A possible source for such additional properties is the Roles Named Authority List[61] maintained by the Publications Office of the EU. Other domain-specific sources for additional properties are the INSPIRE Responsible Party roles[62], the Library of Congress' MARC relators[63] and DataCite's contributor types[64]. To enable the use of such properties, they must be defined as RDF properties with URIs in a well-managed namespace.

A second approach is based on the use of W3C's PROV ontology[65] which provides a powerful mechanism to express a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts. In the context of work on GeoDCAT-AP, a PROV-conformant solution for expressing agent roles was agreed[66]. This solution uses prov:qualifiedAttribution in combination with a dct:type assertion pointing to the code list for Responsible Party Role in the INSPIRE registry. To enable the use of such types, they must be defined with URIs in a well-managed namespace.

Based on the experience gained with the use of domain-specific extensions for additional 'agent roles' in the exchange of information about Datasets, the base DCAT Application

[61] Publications Office of the European Union. Metadata Registry. Authorities. Roles. http://publications.europa.eu/mdr/authority/role/
[62] European Commission. INSPIRE Registry. Responsible Party Role. http://inspire.ec.europa.eu/metadata-codelist/ResponsiblePartyRole/
[63] Library of Congress. MARC Code List for Relators. http://loc.gov/marc/relators/relaterm.html
[64] DataCite Metadata Schema for the Publication and Citation of Research Data, version 3.1. In: https://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf, Appendix 1, table 5.
[65] W3C. PROV-O: The PROV Ontology. W3C Recommendation 30 April 2013. http://www.w3.org/TR/prov-o/
[66] European Commission. Joinup. DCAT application profile for data portals in Europe. GeoDCAT-AP – How to express the different responsible party roles supported in ISO 19115 / INSPIRE. https://joinup.ec.europa.eu/node/141757

Profile may in the future be extended with additional roles that have proven to be useful across domains.

It should be noted that, even if a more expressive approach is used in a particular implementation, the provision of information using dct:publisher for the Catalogue is still mandatory under the rules laid down in the Conformance Statement in section 8, while the provision of information using dct:publisher is strongly recommended for Dataset. The provision of such information using dct:publisher will ensure interoperability with implementations that use the basic approach of DCAT-AP.

10 ACCESSIBILITY AND MULTILINGUAL ASPECTS

Accessibility in the context of this Application Profile is limited to information about the technical format of distributions of datasets. The properties dcat:mediaType and dct:format provide information that can be used to determine what software can be deployed to process the data. The accessibility of the data within the datasets needs to be taken care of by the software that processes the data and is outside of the scope of this Application Profile.

Multilingual aspects related to this Application Profile concern all properties whose contents are expressed as strings (i.e. rdfs:Literal) with human-readable text. Wherever such properties are used, the string values are of one of two types:

- The string is free text. Examples are descriptions and labels. Such text may be translated into several languages.
- The string is an appellation of a 'named entity'. Examples are names of organisations or persons. These names may have parallel versions in other languages but those versions don't need to be literal translations.

Wherever values of properties are expressed with either type of string, the property can be repeated with translations in the case of free text and with parallel versions in case of named entities. For free text, e.g. in the cases of titles, descriptions and keywords, the **language tag** is mandatory.

Language tags to be used with rdfs:Literal are defined by BCP47[67], which allows the use of the "t" extension for text transformations defined in RFC6497[68] with the field "t0"[69] indicating a machine translation.

A language tag will look like: "en-t-es-t0-abcd", which conveys the information that the string is in English, translated from Spanish by machine translation using a tool named "abcd".

For named entities, the language tag is optional and should only be provided if the parallel version of the name is strictly associated with a particular language. For example, the name 'European Union' has parallel versions in all official languages of the union, while a name like 'W3C' is not associated with a particular language and has no parallel versions.

For linking to different language versions of associated web pages (e.g. landing pages) or documentation, a content negotiation[70] mechanism may be used whereby different content is served based on the Accept-Languages indicated by the browser. Using such a mechanism, the link to the page or document can resolve to different language versions of the page or document.

All the occurrences of the property dct:language, which can be repeated if the metadata is provided in multiple languages, MUST have a URI as their object, not a literal string from the ISO 639 code list.

[67] Internet Engineering Task Force (IETF). BCP47. Tags for Identifying Languages. http://tools.ietf.org/html/bcp47
[68] Internet Engineering Task Force (IETF). BCP47 Extension T – Transformed Content. http://tools.ietf.org/html/rfc6497
[69] UNICODE Consortium. CLDR - Unicode Common Locale Data Repository. BCP47, transform_mt.xml. http://unicode.org/cldr/trac/browser/trunk/common/bcp47/transform_mt.xml
[70] Apache Web Server: content negotiation. http://httpd.apache.org/docs/current/content-negotiation.html

How multilingual information is handled in systems, for example in indexing and user interfaces, is outside of the scope of this Application Profile.

11 ACKNOWLEDGEMENTS

This work was elaborated by a Working Group under the ISA programme. The Working Group was co-chaired by **Norbert Hohn** from the Publications Office of the European Union and **Marco Pellegrino** from Eurostat. The ISA Programme of the European Commission was represented by **Vassilios Peristeras** and **Athanasios Karalopoulos**. **Makx Dekkers** was the editor of the specification with important contributions from **Chis Nelson** and **Stefanos Kotoglou**.

The members of the Working Group:

Name	Organisation
Ana Barreda	Spanish National Catalogue.
Anastasia Dimou	Data Science Lab, Ghent University – iMinds.
Agnieszka Zajac	Publications Office of the European Union.
Amandine Masuy	IWEPS, Belgium.
Andrea Perego	Joint Research Centre (JRC).
Andrei Melis	DG CONNECT.
Alan Vask	Marketing and Dissemination Department, Estonia.
Athanasios Karalopoulos	Programme Officer of ISA and responsible for the SEMIC project.
Bert Van Nuffelen	Tenforce, Belgium.
Cristina Miranda	Spanish National Catalogue.
Chris Nelson	Metadata Technology Ltd, UK.
Daniele Rizzi	European Data Portal.
Davide Pesoli	SciamLab, Italy.
Danny Delcambre	Eurostat.
Denis Grofils	Eurostat.
Deirdre Lee	Derilinx, Ireland.
Gregor Boyd	Scottish Government.
Giacomo Gamba	Data dissemination unit of the regional statistics institute of Trento, Italy.
Hubertus Cloodt	Eurostat.
Jakub Klimek	University of Economics in Prague and Ministry or Interior of the Czech Republic.
Jan Dvořák	euroCRIS, the Netherlands.
Jim J. Yang	Agency for Public Management and eGovernment (Difi), Norway.

Makx Dekkers	SEMIC team.
Marco Combetto	Informatica Trentina S.p.A, Italy.
Marco Pellegrino	Eurostat. He is co-chairing the Working Group with Norbert Hohn.
Martial Menard	Eurostat.
Nikolaos Loutas	PwC EU Services, SEMIC team.
Norbert Hohn	Publications Office of the European Union. He is co-chairing the Working Group with Marco Pellegrino.
Paolo Starace	Sciamlab, Italy.
Peter Winstanley	Scottish Government.
Pierre Dumas	Swiss Federal Archives SFA.
Rob Davidson	Office for National Statistics, UK.
Søren Roug	European Environmental Agency.
Stefano Abruzzini	DG CONNECT.
Stefanos Kotoglou	PwC EU Services. SEMIC team.
Uroš Milošević	Tenforce, Belgium.
Valentina Janev	PUPIN, Serbia.
Vassilios Peristeras	Programme Officer of ISA and responsible for the SEMIC project.
Vuk Mijović	PUPIN, Serbia.
Willem Van Gemert	Publications Office of the European Union.

Annex I Quick Reference of Classes and Properties

Class	Class URI	Mandatory properties	Recommended properties	Optional properties	Additional optional properties
Agent	foaf:Agent	foaf:name	dct:type		
Category	skos:Concept	skos:prefLabel			
Category Scheme	skos:ConceptScheme	dct:title			
Catalogue	dcat:Catalog	dcat:dataset dct:description dct:publisher dct:title	foaf:homepage dct:language dct:license dct:issued dcat:themeTaxonomy dct:modified	dct:hasPart dct:isPartOf dcat:record dct:rights dct:spatial	
Catalogue Record	dcat:CatalogRecord	dct:modified foaf:primaryTopic	dct:conformsTo adms:status dct:issued	dct:description dct:language dct:source dct:title	
Checksum	spdx:Checksum	spdx:algorithm spdx:checksumValue			
Dataset	dcat:Dataset	dct:description dct:title	dcat:contactPoint dcat:distribution dcat:keyword dct:publisher dcat:theme	adms:identifier adms:sample adms:versionNotes dcat:landingPage dct:accessRights dct:accrualPeriodicity dct:conformsTo dct:hasVersion dct:isVersionOf dct:identifier dct:issued dct:language dct:modified dct:provenance dct:relation dct:source dct:spatial dct:temporal dct:type foaf:page owl:versionInfo	stat:attribute stat:dimension stat:numSeries dqv:hasQualityAnnotation stat:statMeasure
Distribution	dcat:Distribution	dcat:accessURL	dct:description dct:format dct:license	adms:status dcat:byteSize dcat:downloadURL dcat:mediaType dct:conformsTo dct:issued dct:language dct:modified dct:rights dct:title foaf:page spdx:checksum	dct:type
Document	foaf:Document				
Frequency	dct:Frequency				
Identifier	adms:Identifier	skos:notation			
Kind	vcard:Kind				
Licence Document	dct:LicenseDocument	dct:type			
Licence Type	skos:Concept				
Linguistic System	dct:LinguisticSystem				
Literal	rdfs:Literal				
Location	dct:Location				
Media Type or Extent	dct:MediaTypeOrExtent				
Period Of Time	dct:PeriodOfTime			schema:startDate schema:endDate	
Publisher Type	skos:Concept				
Resource	rdfs:Resource				
Rights Statement	dct:RightsStatement				
Standard	dct:Standard				
Status	skos:Concept				

Annex II Mapping SDMX to DCAT

II.1 Scope

The scope of this section is to describe the mapping of StatDCAT-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 Annex III of this specification), to know which SDMX-ML element or attribute maps to which DCAT-AP class or property.

II.2 Diagrams

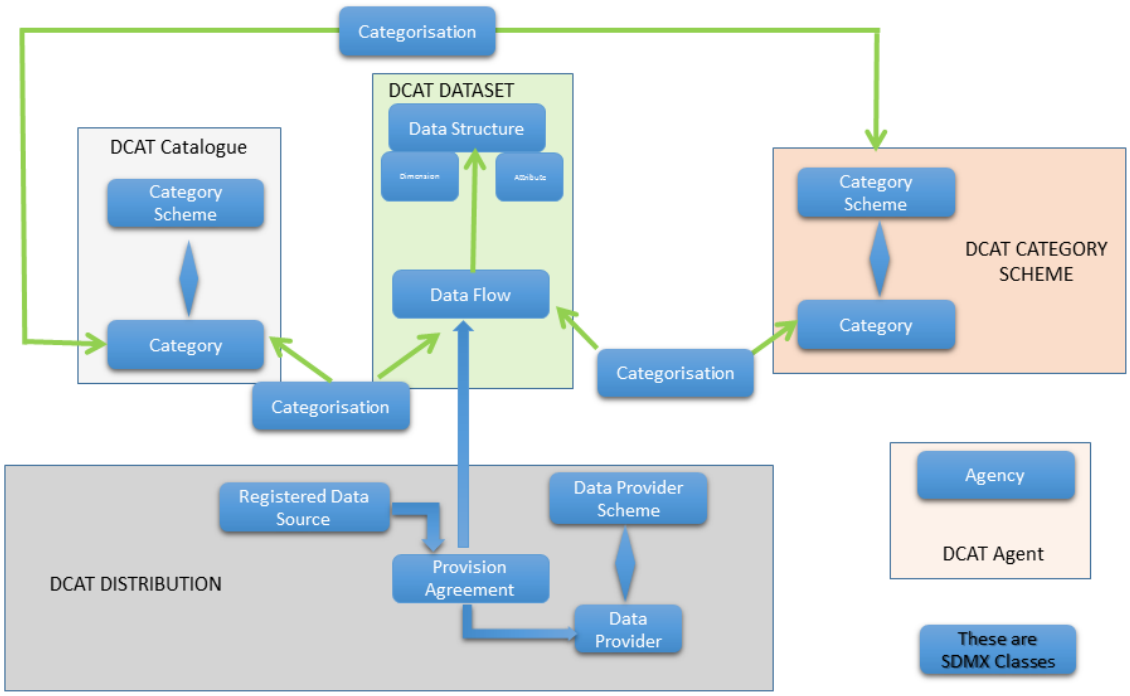


Figure 7: Schematic map of SDMX Classes to DCAT-AP

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 StatDCAT-AP Dataset maps to the SDMX Dataflow.

3) The Dimension and Attribute the StatDCAT-AP Dataset map to the Dimension and Attribute of the SDMX Data Structure

4) 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.

5) 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.

6) 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.

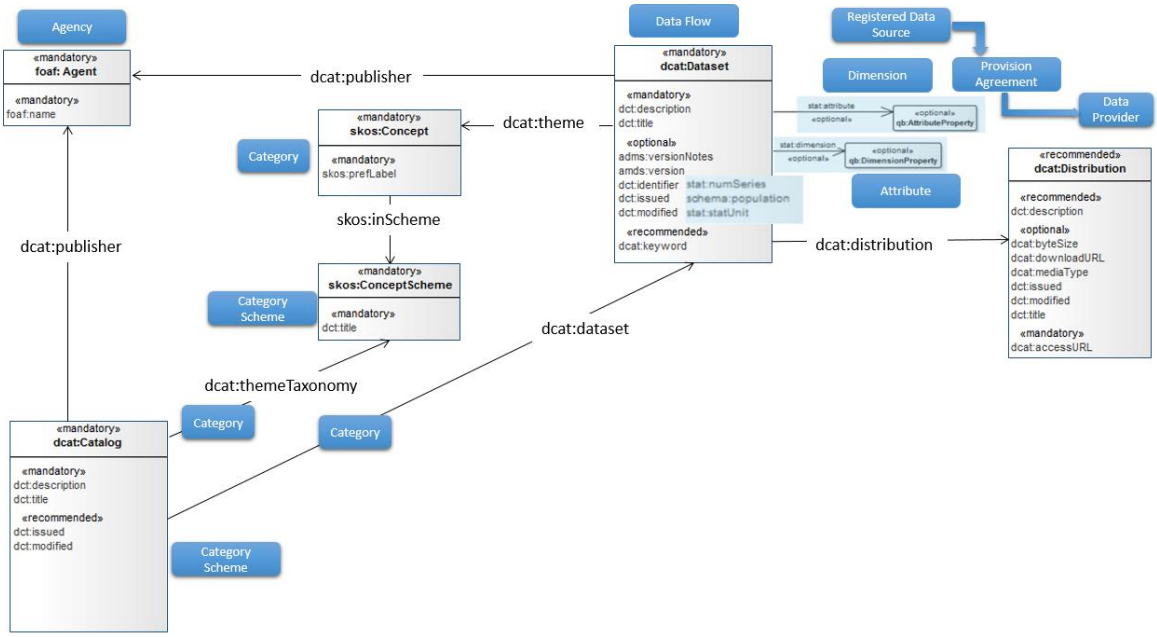


Figure 8: DCAT-AP Model mapped to SDMX Model Classes

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

II.3 Example

II.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.

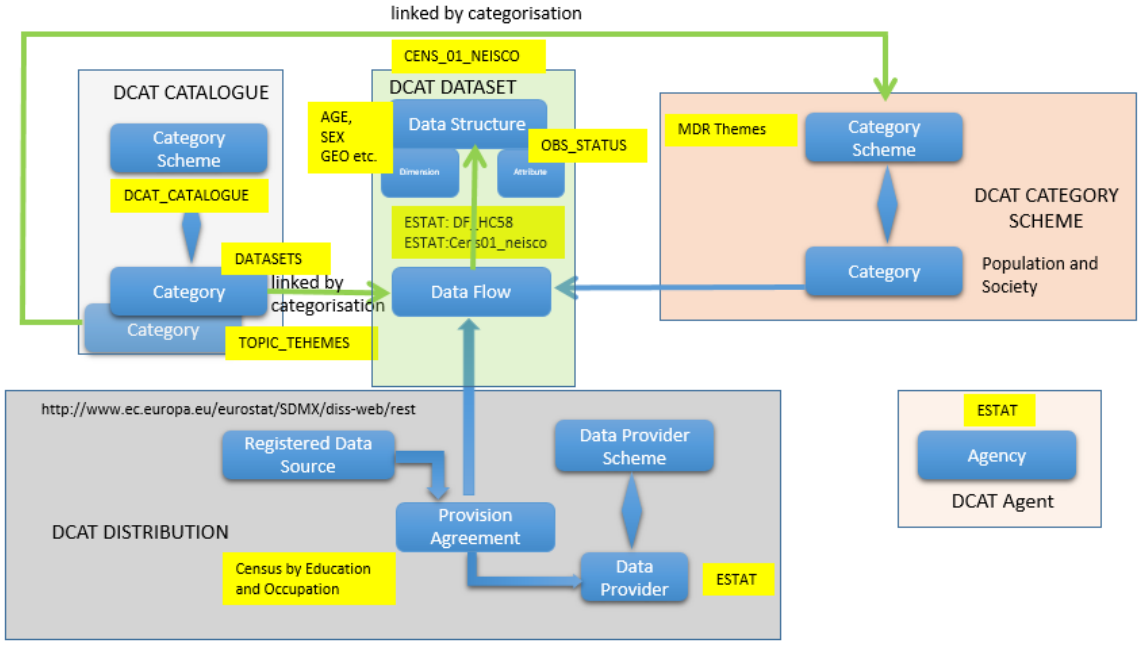


Figure 9: Metadata Used in the Example Mapping

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 Dataflow is linked to the Data Structure *CENS_01_NEISCO* which has a number of Dimensions including Age, Sex, Geography, and an Attribute, Observation Status
4) The SDMX Category Scheme containing the list of has the name *MDR Themes* in the examples.
5) 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.
6) 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*).
7) 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.

II.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:

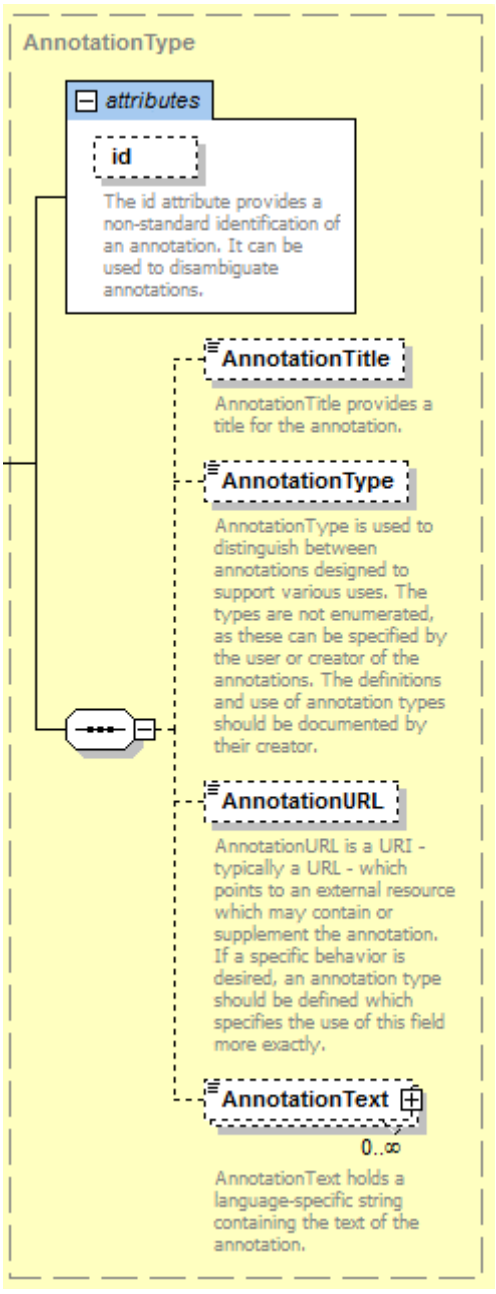


Figure 10: SDMX XML schema specification for Annotation

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)

II.3.3 Explanation of the mapping diagrams

In all cases the mapping is shown between the DCAT-AP property and the location of the property in an SDMX XML message. The table following the mapping diagram lists the DCAT-AP properties and the map to the SDMX XML element or attribute. In all cases the SDMX message is a <Structure> message (i.e. the start tag is the <Structure> element)

e.g. `<mes:Structure`

In the table of properties the additional StatDCAT-AP properties are shown in turquoise.

e.g.

number of data series	stat:numSeries	rdfs:Literal typed as xsd:integer

II.3.4 Data Catalogue

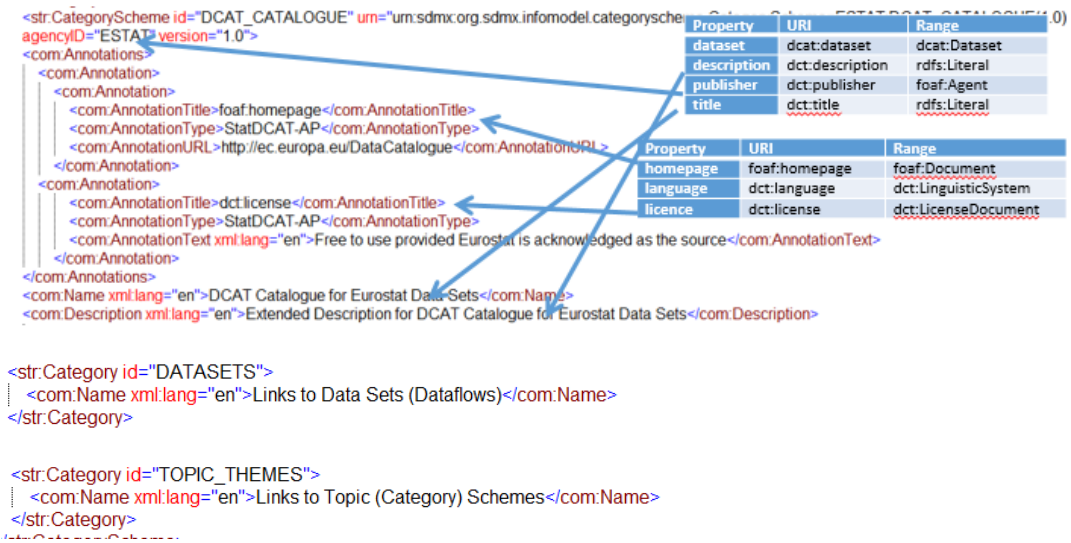


Figure 11: SDMX-DCAT mapping example for the DCAT Catalogue

DCAT-AP Property	SDMX Element or Element.Attribute
dct:description	Description
dct:publisher	CategoryScheme.agencyID
dct:title	Name
foaf:homepage	AnnotationTitle
dct:license	AnnotationTitle

II.3.5 Linking to Categories using Categorisations

Schematic

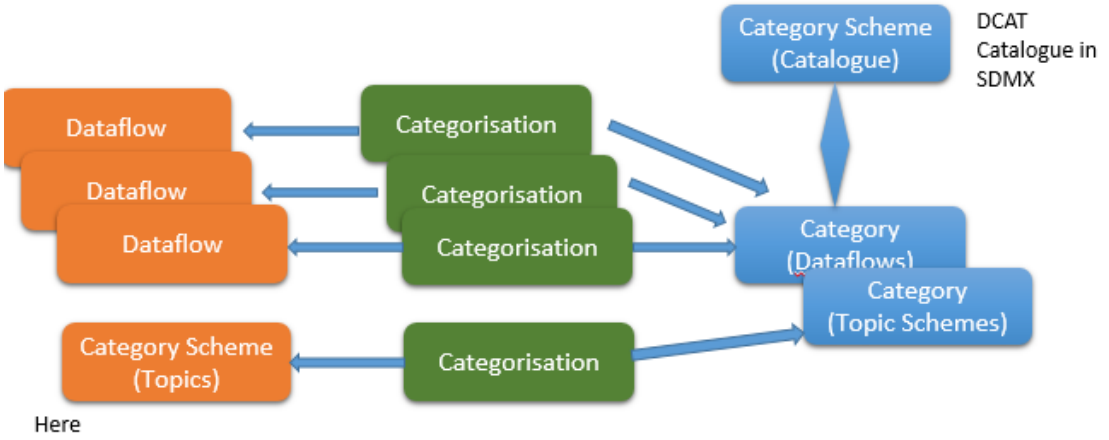


Figure 12: Schematic showing linking of SDMX Categories to other SDMX objects

Example

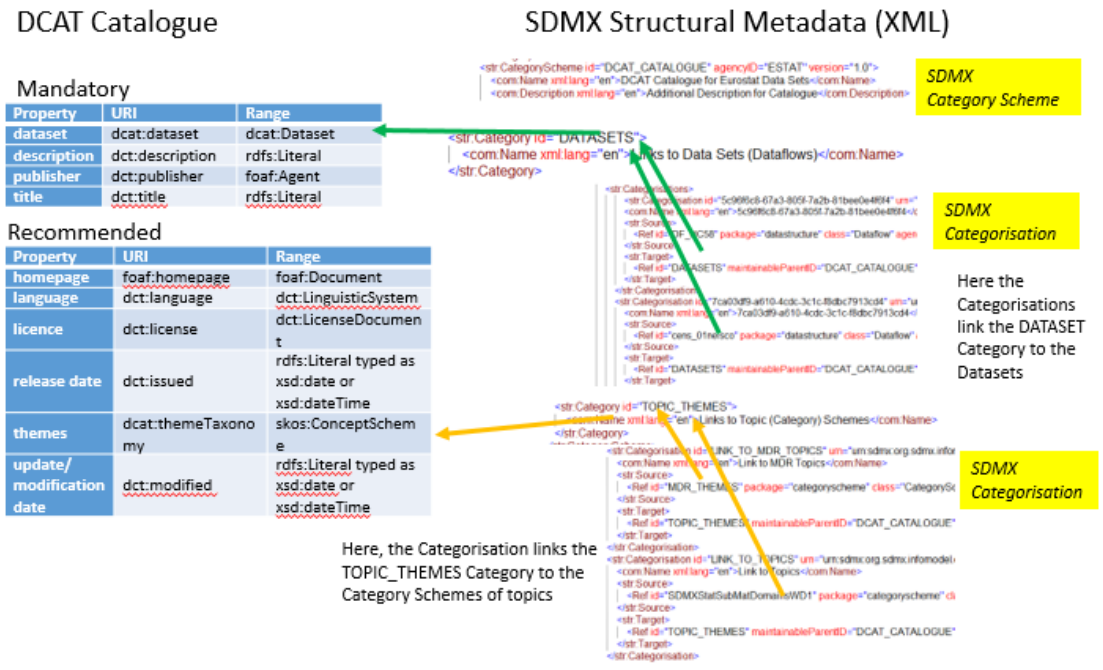


Figure 13: Linking Catalogue to DCAT Datasets and Category (Topic) Scheme

DCAT-AP Property	SDMX Element or Element.Attribute
dcat:dataset	Source/Ref.id (id of the Dataflow Source/Ref.agencyId (agency of the Dataflow) Source/Ref.version (version of the Dataflow) Target/Ref.id (id of the Category) Target/maintainableParentId (id of the Category Scheme that is the DCAT-AP Catalogue) Target/agencyId (agency of the Category Scheme) Target/version (version of the Category Scheme)
dcat:themeTaxonomy	Source/Ref.id (id of the Dataflow) Source/Ref.agencyId (agency of the Dataflow) Source/Ref.version (version of the Dataflow) Target/Ref.id (id of the Category) Target/maintainableParentId (id of the Category Scheme that is the DCAT-AP Category Scheme) Target/agencyId (agency of the Category Scheme) Target/version (version of the Category Scheme)

II.3.6 Dataset

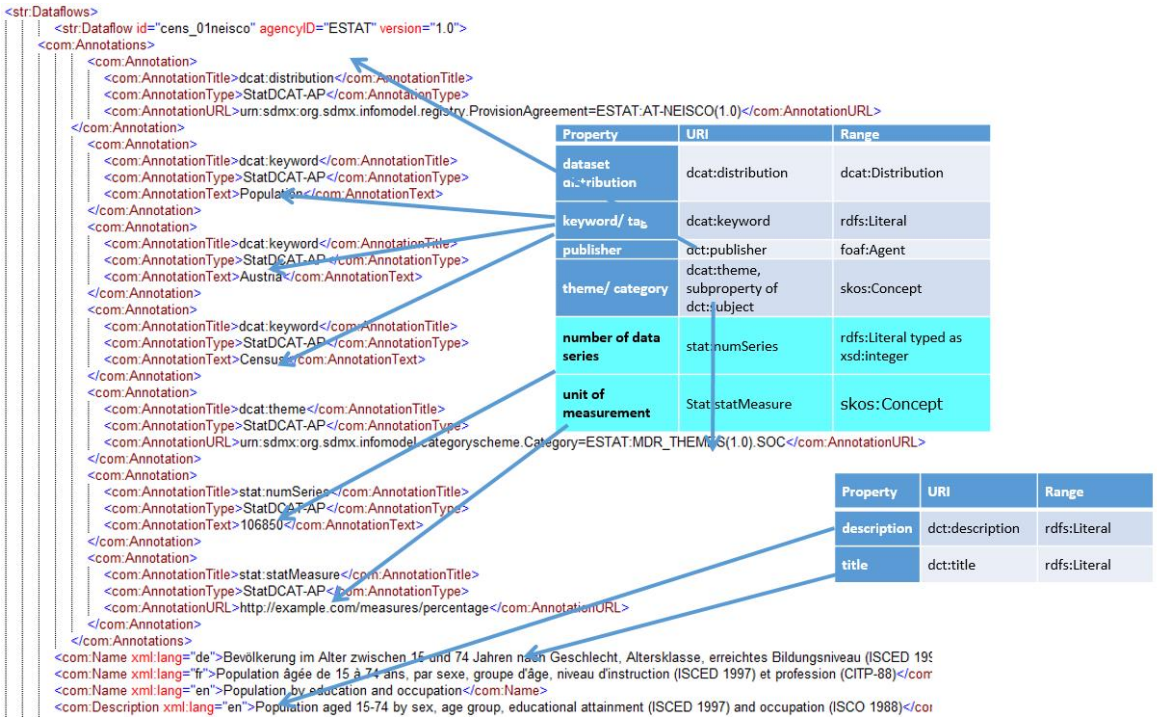


Figure 14: SDMX to DCAT mapping example for the StatDCAT-AP Dataset

DCAT-AP Property	SDMX Element or Element.Attribute
dcat:distribution	AnnotationURL
dcat:keyword	AnnotatationText
dct:publisher	Dataflow.agencyID
dcat:theme	AnnotationURL
stat:numSeries	AnnotationText
stat:statMeasure	AnnotationURL
dct:description	Description
dct:title	Name

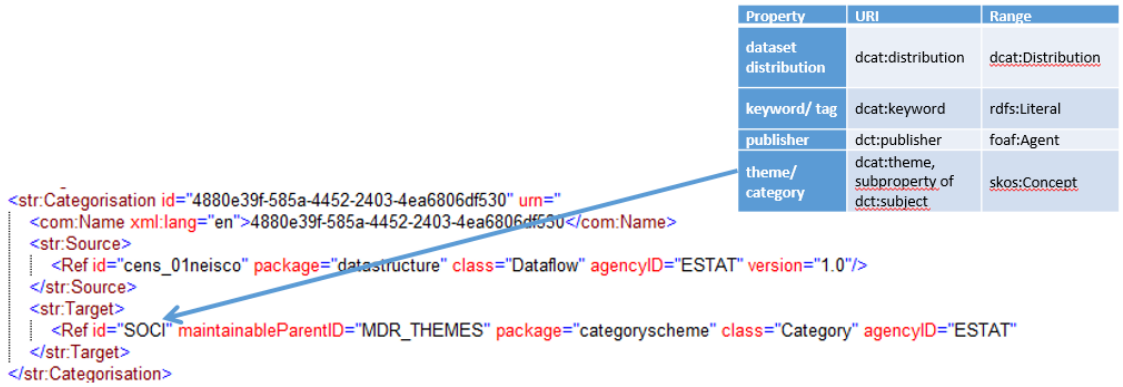


Figure 15: Linking a Dataflow to the SDMX Category (Topic)

This links the Dataflow cens_01neisco version 1.0 maintained by ESTAT to the Category SOCI in the Category Scheme representing the MDR Themes (MDR_THEMES). This (SDMX) Category is the map to the dcat:theme.

II.3.7 Dimension Property and Attribute Property

The URL must resolve to a qb:dimension or qb:attribute.

II.3.8 Quality Annotation

If required as SDMX structural metadata this will be an Annotation on the Dataflow

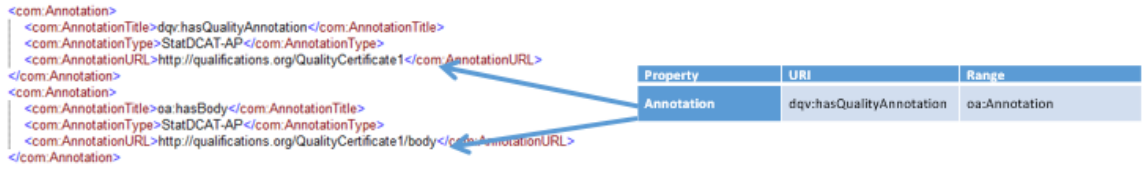


Figure 16: SDMX to DCAT mapping example for the StatDCAT-AP Annotation

II.3.9 Distribution

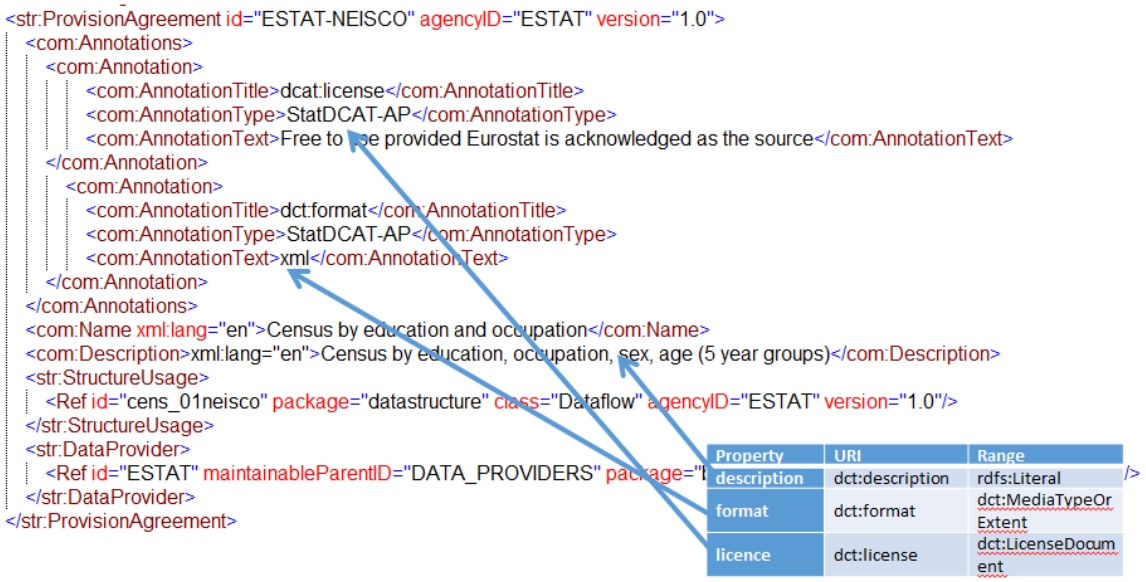


Figure 17: Linking a Distribution to the SDMX Provision Agreement

DCAT-AP Property	SDMX Element or Eelement.Attribute
dct:description	Description
dct:format	AnnotationText
dct:license	AnnotationText

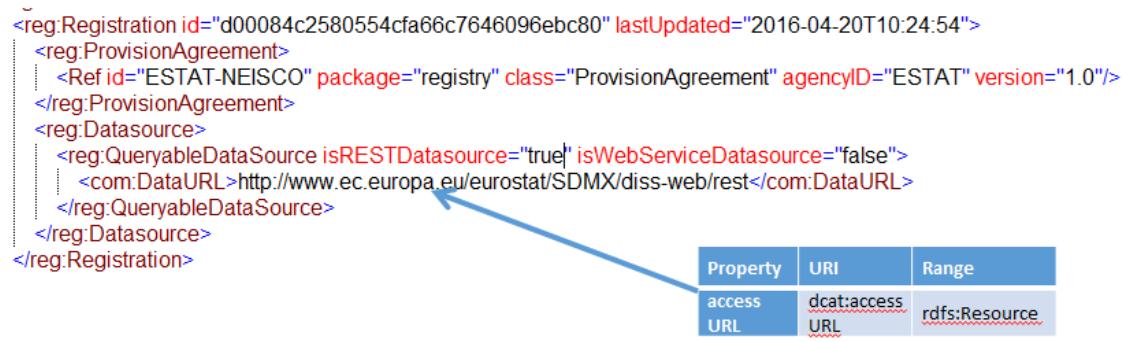


Figure 18: Linking a Distribution (accessURL) to the SDMX Provision Agreement

DCAT-AP Property	SDMX Element or Eelement.Attribute
dcat:accessURL	QueryableDataSource/DataURL

II.3.10 Agent

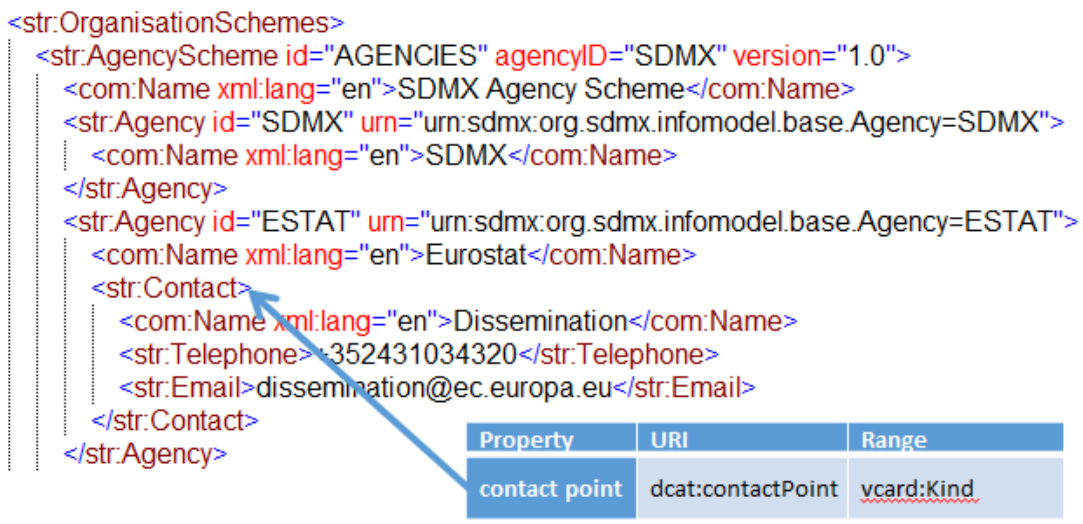


Figure 19:Linking an Agent to the SDMX Agency

DCAT-AP Property	SDMX Element or Eelement.Attribute
dcat:contactPoint	Contact/Name Contact/Telephone Contact/Email

II.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 (this contains the DCAT-AP property).

Annex III SDMX-BASED TRANSFORMATION MECHANISM

III.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, it is the intention that 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.

III.2 Transformation mechanism

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

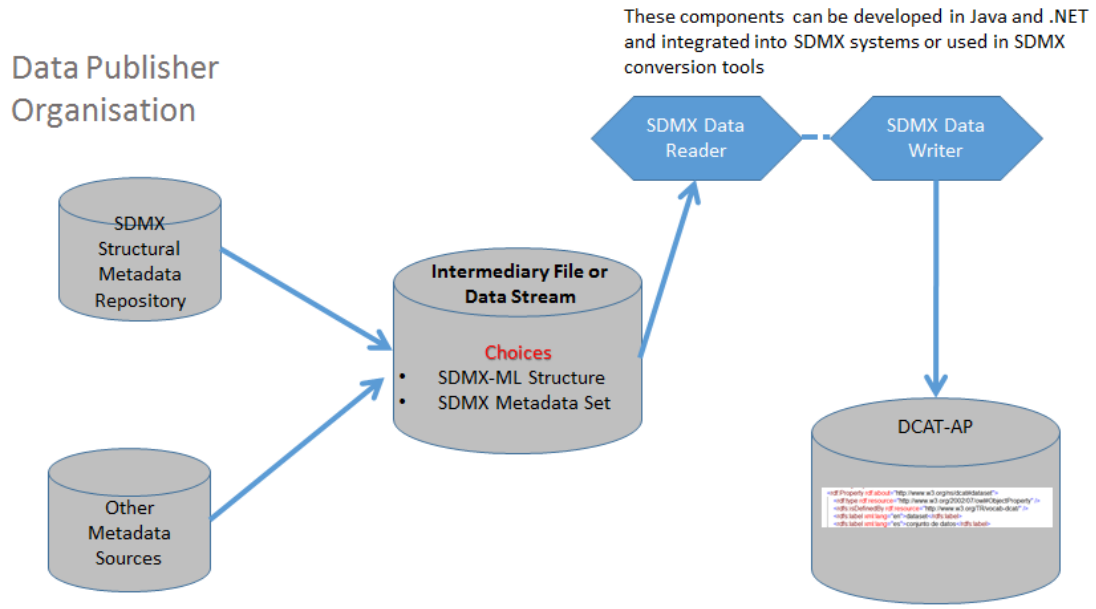


Figure 20: Diagram of the flow of metadata though the Intermediary Mechanism

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.

III.3 Transformation input formats

III.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.

III.3.2 SDMX Structural Metadata

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

However, there is one difference between the mapping given in Annex II 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 Annex II:

```
<reg:Registration id="d00084c2580554cfa66c7646096ebc80" lastUpdated="2016-04-20T10:24:54">
  <reg:ProvisionAgreement>
    <Ref id="ESTAT-NEISCO" package="registry" class="ProvisionAgreement" agencyID="ESTAT" version="1.0"/>
  </reg:ProvisionAgreement>
  <reg:Datasource>
    <reg:QueryableDataSource isRESTDatasource="true" isWebServiceDatasource="false">
      <com:DataURL>http://www.ec.europa.eu/eurostat/SDMX/diss-web/rest</com:DataURL>
    </reg:QueryableDataSource>
  </reg:Datasource>
</reg:Registration>
```

Property	URI	Range
access URL	dcat:access URL	rdfs:Resource

Figure 21: From Section 9 - Linking a Distribution (accessURL) to the SDMX Provision Agreement

Using the Transformation Mechanism this is output as:

```
<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>
```

Property	URI	Range
access URL	dcat:access URL	rdfs:Resource

Figure 22: 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 23: Example Provision Agreement for DCAT-AP Distribution

The full example Structure Message is shown in Annex IV.

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.

III.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.

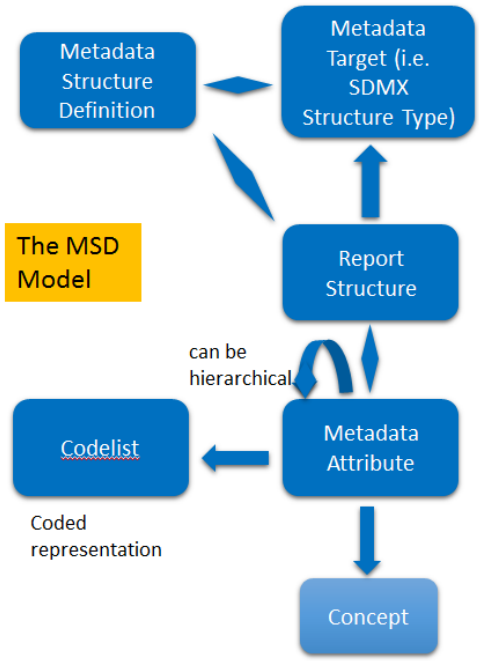


Figure 24: Schematic diagram of the SDMX Metadata Structure Definition

A schematic of the Metadata Set is shown below.

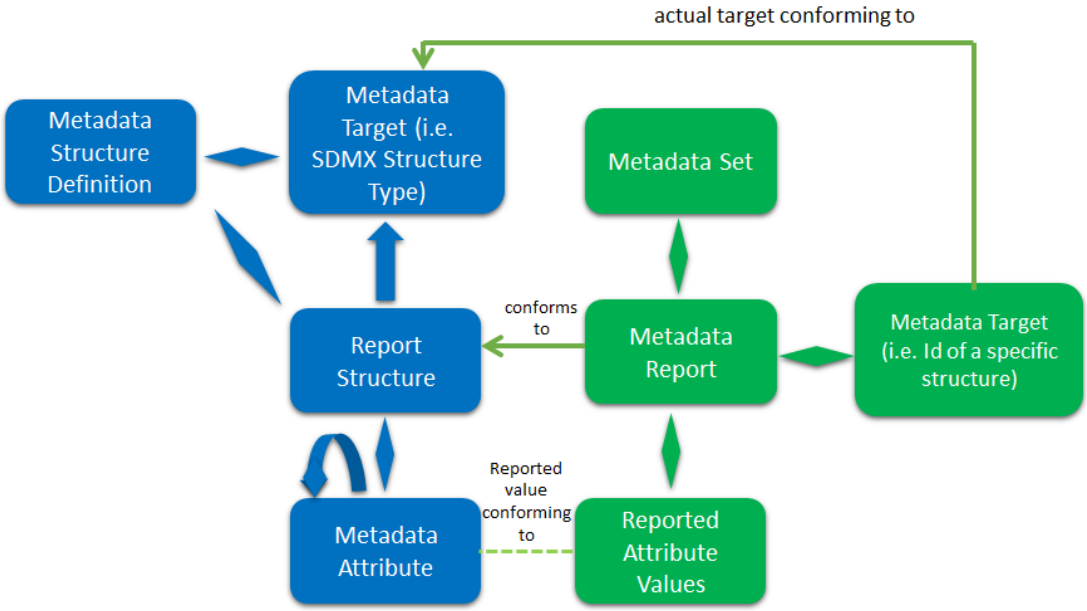


Figure 25: Schematic diagram of the SDMX Metadata Set

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-AP Catalogue.

Example MSD

Note that this MSD is not finalised. At the moment is contains the Mandatory and Recommended properties of DVAT-AP, and the extensions added by StatDCAT-AP.

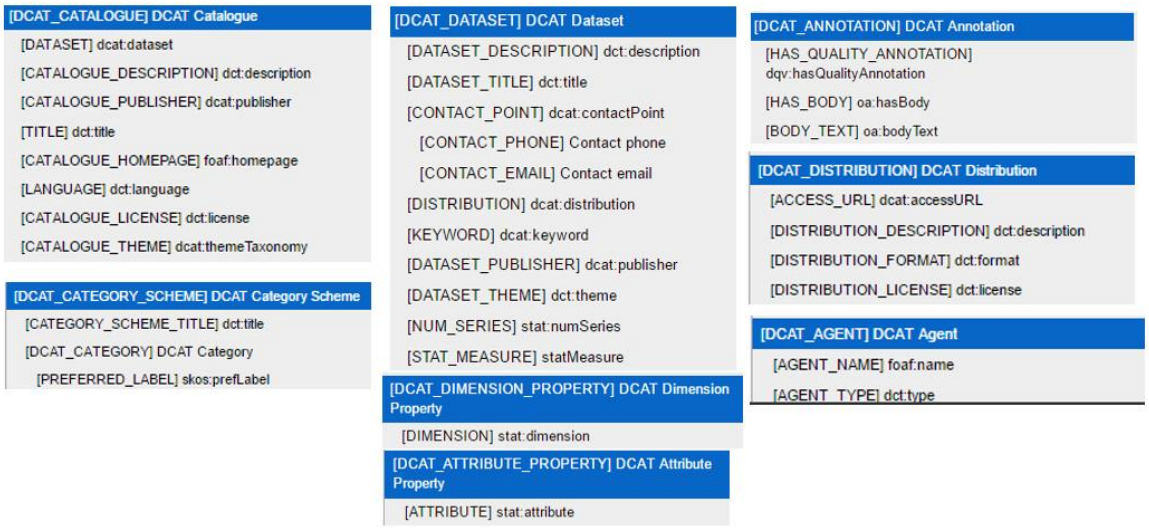


Figure 26: Metadata Attributes in the DCAT-AP MSD

Each DCAT-AP class and StatDCAT-AP class are top level Metadata Attributes in the MSD. The properties of the class are the child Metadata Attributes. 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.

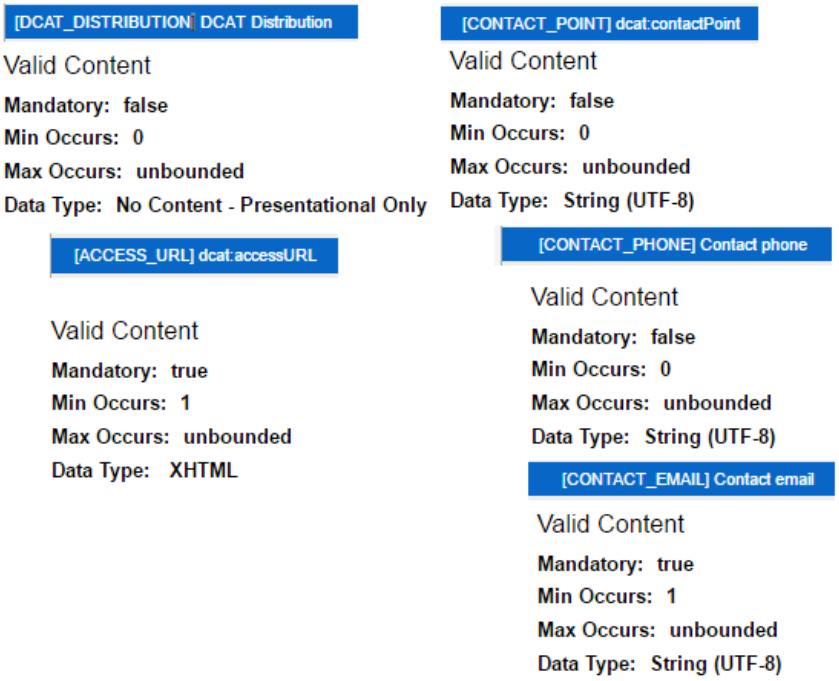


Figure 27: Example Metadata Attribute Specification

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 Set Report shows how the DCAT-AP metadata are represented in a Metadata Set structured according to the MSD.

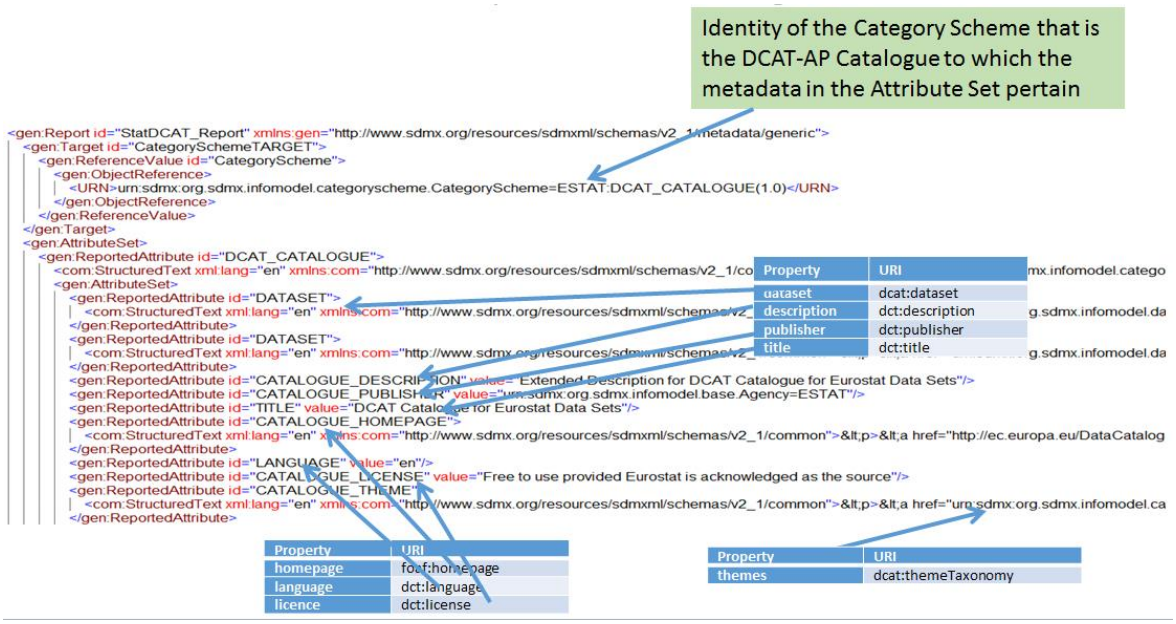


Figure 28: SDMX catalogue metadata pertaining to the DCAT-AP Catalogue

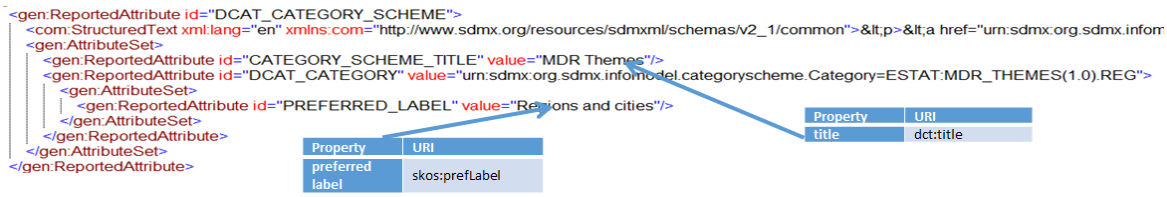


Figure 29: SDMX category scheme metadata pertaining to the DCAT-AP Catalogue

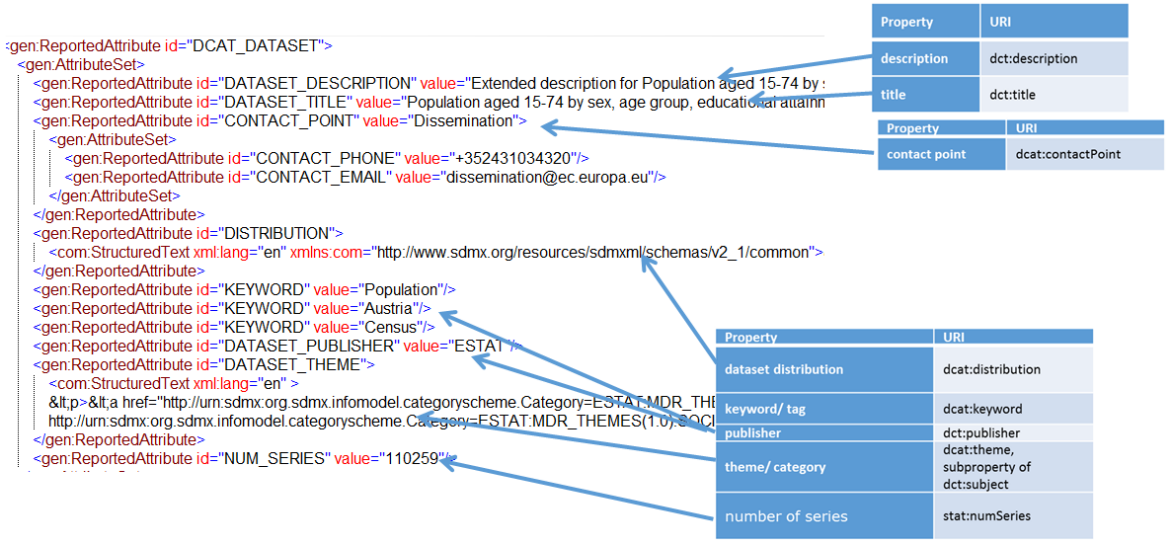


Figure 30: SDMX dataset metadata pertaining to the DCAT-AP Catalogue including StatDCAT-AP extensions to the Dataset.

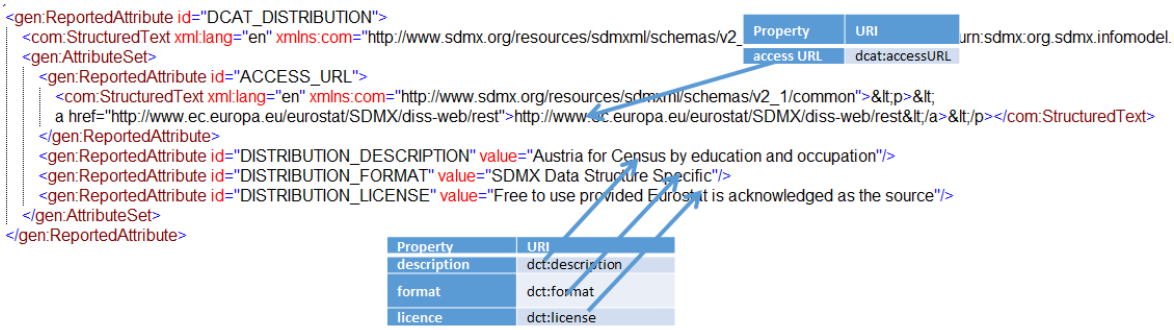


Figure 31: SDMX distribution metadata pertaining to the DCAT-AP Catalogue including StatDCAT-AP extensions to the Distribution

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.

III.4 Advantages and disadvantages of the two transformation formats

III.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.

III.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 Metadata Set Report 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

III.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 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.

Annex IV SDMX FILES USED FOR THE EXAMPLES

IV.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="cens_01neisco" agencyID="ESTAT" 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:Annotation>
              <com:AnnotationTitle>stat:numSeries</com:AnnotationTitle>
              <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
              <com:AnnotationText>106850</com:AnnotationText>
            </com:Annotation>
            <com:Annotation>
              <com:AnnotationTitle>stat:statMeasure</com:AnnotationTitle>
              <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
              <com:AnnotationURL>http://example.com/measures/percentage</com:AnnotationURL>
            </com:Annotation>
            <com:Annotation>
              <com:AnnotationTitle>dqv:hasQualityAnnotation</com:AnnotationTitle>
              <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
              <com:AnnotationURL>http://qualifications.org/QualityCertificate1</com:AnnotationURL>
            </com:Annotation>
            <com:Annotation>
              <com:AnnotationTitle>oa:hasBody</com:AnnotationTitle>
              <com:AnnotationType>StatDCAT-AP</com:AnnotationType>
              <com:AnnotationURL>http://qualifications.org/QualityCertificate1/body</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>
```

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>
```

IV.2 SDMX Metadata Set

IV.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">&lt;p&gt;&lt;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)&lt;/a>&lt;/
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">&lt;p&gt;&lt;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&lt;/a>&lt;/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">&lt;p&gt;&lt;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)&lt;/a>&lt;/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">&lt;p&gt;&lt;a
href="http://ec.europa.eu/DataCatalogue">http://ec.europa.eu/DataCatalogue&lt;/a>&lt;/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">&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>&lt;/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>&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

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

http://urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).SOCI

stat:numSeries

110259

statMeasure

http://example.com/measures/percentage

<gen:ReportedAttribute id="DCAT_DATASET">
<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"><p>urn:sdmx:org.sdmx.infomodel.registry.ProvisionAgreement=ESTAT:AT-NEISCO(1.0)</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">
<com:StructuredText xml:lang="en" xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"><p>http://urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR_THEMES(1.0).SOCI</p></com:StructuredText>
</gen:ReportedAttribute>
<gen:ReportedAttribute id="NUM_SERIES" value="110259"/>
<gen:ReportedAttribute id="STAT_MEASURE">
<com:StructuredText xml:lang="en" xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"><p>http://example.com/measures/percentage</p></com:StructuredText>
</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">&lt;p>&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>&lt;/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">&lt;p>&lt;a
href="http://localhost:8080/FusionRegistry/ws/rest">http://localhost:8080/FusionRegistry/ws/rest&lt;/a>&lt;/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">&lt;p>&lt;a
href="urn:sdmx:org.sdmx.infomodel.base.Agency=ESTAT">urn:sdmx:org.sdmx.infomodel.base.Agency=EST
AT&lt;/a>&lt;/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>
```