SC353 DI07171

D02.02.03: StatDCAT-AP – DCAT Application Profile for description of statistical datasets, Final draft

Document Metadata

|  |  |
| --- | --- |
| Date | 2016-11-25 |
| Status | Final |
| Version | 0.06 |
| Authors | Makx Dekkers – AMI Consult  Chris Nelson – Metadata Technologies  Stefanos Kotoglou – PwC EU Services  Florian Barthelemy – 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:**

The views expressed in this report are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission.

The European Commission does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof.

Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission.

All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.

Table of Contents

[1 Introduction 1](#_Toc467832128)

[1.1 Background 1](#_Toc467832129)

[1.2 Objectives 2](#_Toc467832130)

[1.3 Roadmap 3](#_Toc467832131)

[1.4 Structure of this document 4](#_Toc467832132)

[2 Terminology used in this document 5](#_Toc467832133)

[3 Related work 7](#_Toc467832134)

[3.1 Statistical data and metadata initiatives 7](#_Toc467832135)

[3.1.1 Eurostat and EU Publications Office collaboration 7](#_Toc467832136)

[3.1.2 SDMX 7](#_Toc467832137)

[3.1.3 ESMS 9](#_Toc467832138)

[3.2 Open Data standards and application profiles 10](#_Toc467832139)

[3.2.1 W3C DCAT 10](#_Toc467832140)

[3.2.2 DCAT-AP for open data portals in Europe 11](#_Toc467832141)

[3.2.3 GeoDCAT-AP 13](#_Toc467832142)

[3.2.4 The Data Cube Vocabulary 14](#_Toc467832143)

[4 Use cases 15](#_Toc467832144)

[4.1 Improve discoverability of statistical datasets on open data portals 15](#_Toc467832145)

[4.2 Federation of open data portals 15](#_Toc467832146)

[5 Methodology 16](#_Toc467832147)

[5.1 ISA Core Vocabulary process and methodology 16](#_Toc467832148)

[5.2 Analysis and decision framework 16](#_Toc467832149)

[5.3 Stakeholders 16](#_Toc467832150)

[5.4 Time plan 17](#_Toc467832151)

[6 The StatDCAT-AP data model 18](#_Toc467832152)

[6.1 Informal description 18](#_Toc467832153)

[6.2 Extensions and specific usage for description of statistical datasets 19](#_Toc467832154)

[6.2.1 Dimensions and attributes 19](#_Toc467832155)

[6.2.2 Quality aspects 20](#_Toc467832156)

[6.2.3 Visualisation 21](#_Toc467832157)

[6.2.4 Number of data series 21](#_Toc467832158)

[6.2.5 Unit of measurement 22](#_Toc467832159)

[6.2.6 Specifying the length of time series 22](#_Toc467832160)

[6.3 Overview of the model 23](#_Toc467832161)

[6.4 Namespaces 24](#_Toc467832162)

[6.5 UML Class diagram 25](#_Toc467832163)

[6.6 Description of classes 26](#_Toc467832164)

[6.6.1 Mandatory Classes 26](#_Toc467832165)

[6.6.2 Recommended Classes 26](#_Toc467832166)

[6.6.3 Optional Classes 26](#_Toc467832167)

[6.7 Description of properties per class 28](#_Toc467832168)

[6.7.1 Catalogue 28](#_Toc467832169)

[6.7.2 Catalogue Record 28](#_Toc467832170)

[6.7.3 Dataset 29](#_Toc467832171)

[6.7.4 Distribution 32](#_Toc467832172)

[6.7.5 Agent 33](#_Toc467832173)

[6.7.6 Category Scheme 33](#_Toc467832174)

[6.7.7 Category 33](#_Toc467832175)

[6.7.8 Checksum 33](#_Toc467832176)

[6.7.9 Identifier 33](#_Toc467832177)

[6.7.10 Licence Document 34](#_Toc467832178)

[6.7.11 Period of Time 34](#_Toc467832179)

[6.8 Controlled vocabularies 34](#_Toc467832180)

[6.8.1 Requirements for controlled vocabularies 34](#_Toc467832181)

[6.8.2 Controlled vocabularies to be used 35](#_Toc467832182)

[6.8.3 Mapping Eurostat theme vocabulary to the MDR data themes vocabulary 36](#_Toc467832183)

[6.8.4 Other controlled vocabularies 36](#_Toc467832184)

[6.8.5 Licence vocabularies 37](#_Toc467832185)

[7 Mapping and Extraction approaches 38](#_Toc467832186)

[8 Conformance statement 39](#_Toc467832187)

[8.1 Provider requirements 39](#_Toc467832188)

[8.2 Receiver requirements 39](#_Toc467832189)

[9 Agent roles 40](#_Toc467832190)

[10 Date and Time 42](#_Toc467832191)

[11 Accessibility and Multilingual Aspects 43](#_Toc467832192)

[12 Acknowledgements 45](#_Toc467832193)

[Annex I Quick Reference of Classes and Properties 47](#_Toc467832194)

[Annex II Mapping SDMX to DCAT 49](#_Toc467832195)

[II.1 Scope 49](#_Toc467832196)

[II.2 Diagrams 49](#_Toc467832197)

[II.3 Example 51](#_Toc467832198)

[II.3.1 Introduction 51](#_Toc467832199)

[II.3.2 SDMX Annotations 52](#_Toc467832200)

[II.3.3 Explanation of the mapping diagrams 54](#_Toc467832201)

[II.3.4 Data Catalogue 54](#_Toc467832202)

[II.3.5 Linking to Categories using Categorisations 55](#_Toc467832203)

[II.3.6 Dataset 56](#_Toc467832204)

[II.3.7 Dimension Property and Attribute Property 57](#_Toc467832205)

[II.3.8 Quality Annotation 57](#_Toc467832206)

[II.3.9 Distribution 58](#_Toc467832207)

[II.3.10 Agent 59](#_Toc467832208)

[II.4 Summary 59](#_Toc467832209)

[Annex III SDMX-based Transformation Mechanism 60](#_Toc467832210)

[III.1 Scope of this section 60](#_Toc467832211)

[III.2 Transformation mechanism 60](#_Toc467832212)

[III.3 Transformation input formats 61](#_Toc467832213)

[III.3.1 Choice of mechanisms 61](#_Toc467832214)

[III.3.2 SDMX Structural Metadata 61](#_Toc467832215)

[III.3.3 SDMX Metadata Set 63](#_Toc467832216)

[III.4 Advantages and disadvantages of the two transformation formats 67](#_Toc467832217)

[III.4.1 SDMX Structure Message 67](#_Toc467832218)

[III.4.2 SDMX Metadata Set 68](#_Toc467832219)

[III.5 Summary 68](#_Toc467832220)

[Annex IV SDMX Files used for the examples 69](#_Toc467832221)

[IV.1 SDMX Structural Metadata 69](#_Toc467832222)

[IV.2 SDMX Metadata Set 73](#_Toc467832223)

[IV.2.1 Content 73](#_Toc467832224)

[Annex V Examples of StatDCAT-AP descriptions of Data Cube DataSets 78](#_Toc467832225)

[V.1 RDF Example 1 78](#_Toc467832226)

[V.2 RDF Example 2 79](#_Toc467832227)

List of Figures

[Figure 1: SDMX Main Components 8](#_Toc467832228)

[Figure 2: SDMX Information Model: Schematic View 8](#_Toc467832229)

[Figure 3: Fragment of ESMS specification 9](#_Toc467832230)

[Figure 4: DCAT schematic data model 11](#_Toc467832231)

[Figure 5: DCAT-AP Data Model 12](#_Toc467832232)

[Figure 6: Data Cube vocabulary overview of key terms and relationships 14](#_Toc467832233)

[Figure 7: Schematic map of SDMX Classes to DCAT-AP 49](#_Toc467832234)

[Figure 8: DCAT-AP Model mapped to SDMX Model Classes 51](#_Toc467832235)

[Figure 9: Metadata Used in the Example Mapping 51](#_Toc467832236)

[Figure 10: SDMX XML schema specification for Annotation 53](#_Toc467832237)

[Figure 11: SDMX-DCAT mapping example for the DCAT Catalogue 54](#_Toc467832238)

[Figure 12: Schematic showing linking of SDMX Categories to other SDMX objects 55](#_Toc467832239)

[Figure 13: Linking Catalogue to DCAT Datasets and Category (Topic) Scheme 55](#_Toc467832240)

[Figure 14: SDMX to DCAT mapping example for the StatDCAT-AP Dataset 56](#_Toc467832241)

[Figure 15: Linking a Dataflow to the SDMX Category (Topic) 57](#_Toc467832242)

[Figure 16: SDMX to DCAT mapping example for the StatDCAT-AP Annotation 57](#_Toc467832243)

[Figure 17: Linking a Distribution to the SDMX Provision Agreement 58](#_Toc467832244)

[Figure 18: Linking a Distribution (accessURL) to the SDMX Provision Agreement 58](#_Toc467832245)

[Figure 19:Linking an Agent to the SDMX Agency 59](#_Toc467832246)

[Figure 20: Diagram of the flow of metadata though the Intermediary Mechanism 60](#_Toc467832247)

[Figure 21: From Section 9 - Linking a Distribution (accessURL) to the SDMX Provision Agreement 62](#_Toc467832248)

[Figure 22: Transformation format - Linking a Distribution (accessURL) to the SDMX Provision Agreement 62](#_Toc467832249)

[Figure 23: Example Provision Agreement for DCAT-AP Distribution 62](#_Toc467832250)

[Figure 24: Schematic diagram of the SDMX Metadata Structure Definition 63](#_Toc467832251)

[Figure 25: Schematic diagram of the SDMX Metadata Set 64](#_Toc467832252)

[Figure 26: Metadata Attributes in the DCAT-AP MSD 64](#_Toc467832253)

[Figure 27: Example Metadata Attribute Specification 65](#_Toc467832254)

[Figure 28: SDMX catalogue metadata pertaining to the DCAT-AP Catalogue 66](#_Toc467832255)

[Figure 29: SDMX category scheme metadata pertaining to the DCAT-AP Catalogue 66](#_Toc467832256)

[Figure 30: SDMX dataset metadata pertaining to the DCAT-AP Catalogue including StatDCAT-AP extensions to the Dataset. 66](#_Toc467832257)

[Figure 31: SDMX distribution metadata pertaining to the DCAT-AP Catalogue including StatDCAT-AP extensions to the Distribution 67](#_Toc467832258)

# Introduction

## 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 should 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]](#footnote-1) and in its EU implementation[[2]](#footnote-2). This statement is confirmed in the Commission’s Notice 2014/C 240/01[[3]](#footnote-3), elaborating the results of the online consultation launched by the Commission in August 2013 on the revision to the PSI Directive[[4]](#footnote-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]](#footnote-5) has started operation in November 2015. Statistical data is of great interest for all 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]](#footnote-6), developed under the aegis of the European Commission’s ISA programme[[7]](#footnote-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 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]](#footnote-8) of the work done in 2015 is available from the European Commission’s ISA programme.

## 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:

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.

## 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.

## 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 explains the expression of date and time.

Section 11 includes clarifications regarding accessibility and multilingual aspects.

Section 12 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.

# 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]](#footnote-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]](#footnote-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. |
| **MDR – Metadata Registry** | The Metadata Registry[[11]](#footnote-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]](#footnote-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]](#footnote-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]](#footnote-14) specifying the syntax and semantics of formalized information for location and access of resources via the Internet. |

# Related work

## Statistical data and metadata initiatives

### Eurostat and EU Publications Office collaboration

In the context of the European Union Open Data Portal (EU ODP)[[15]](#footnote-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]](#footnote-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.

### 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.

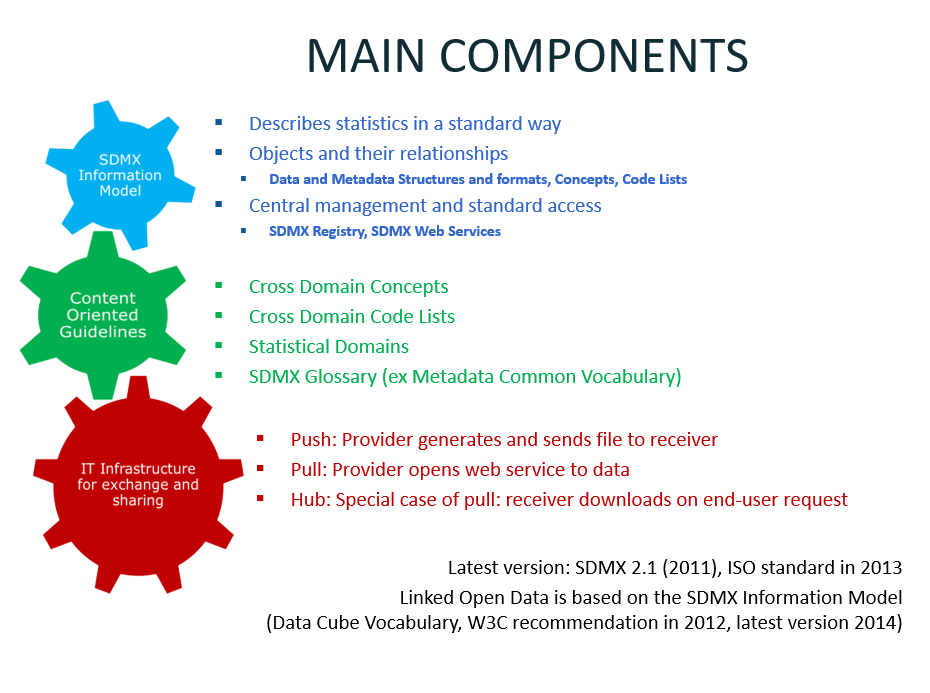


Figure : SDMX Main Components

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

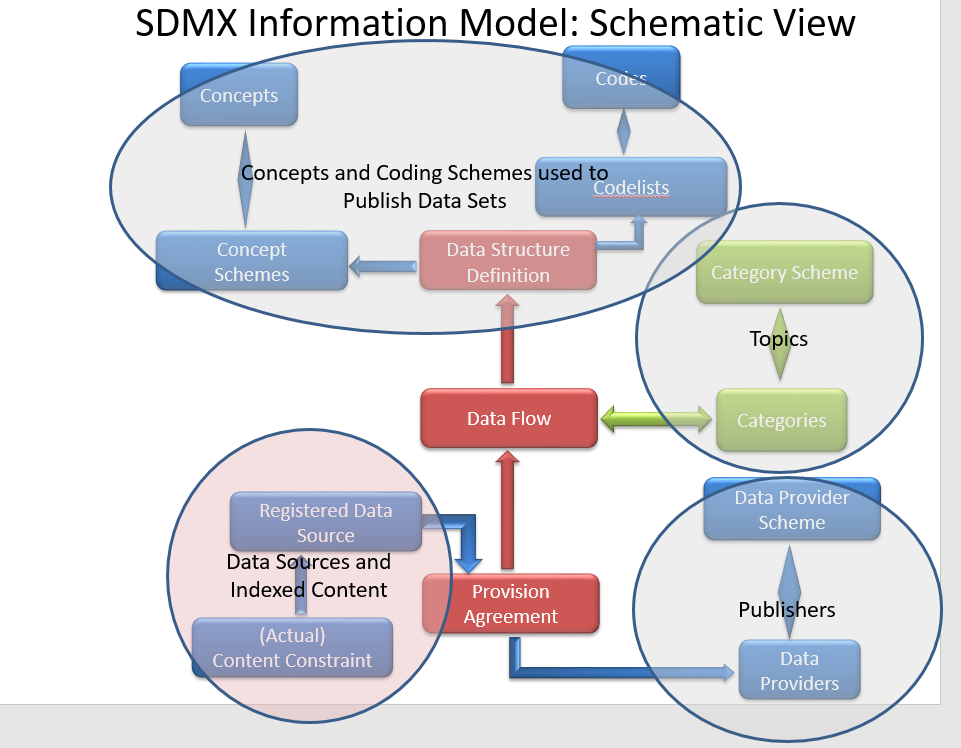


Figure : SDMX Information Model: Schematic View

### ESMS

The Euro SDMX Metadata Structure (ESMS)[[17]](#footnote-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.

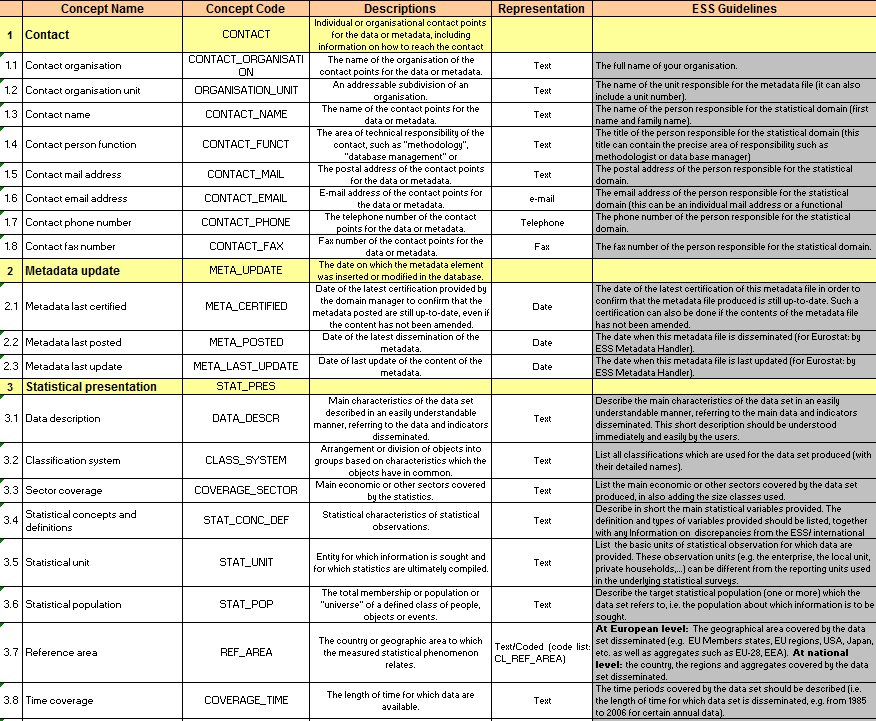


Figure : Fragment of ESMS specification

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.

## Open Data standards and application profiles

### 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]](#footnote-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]](#footnote-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.

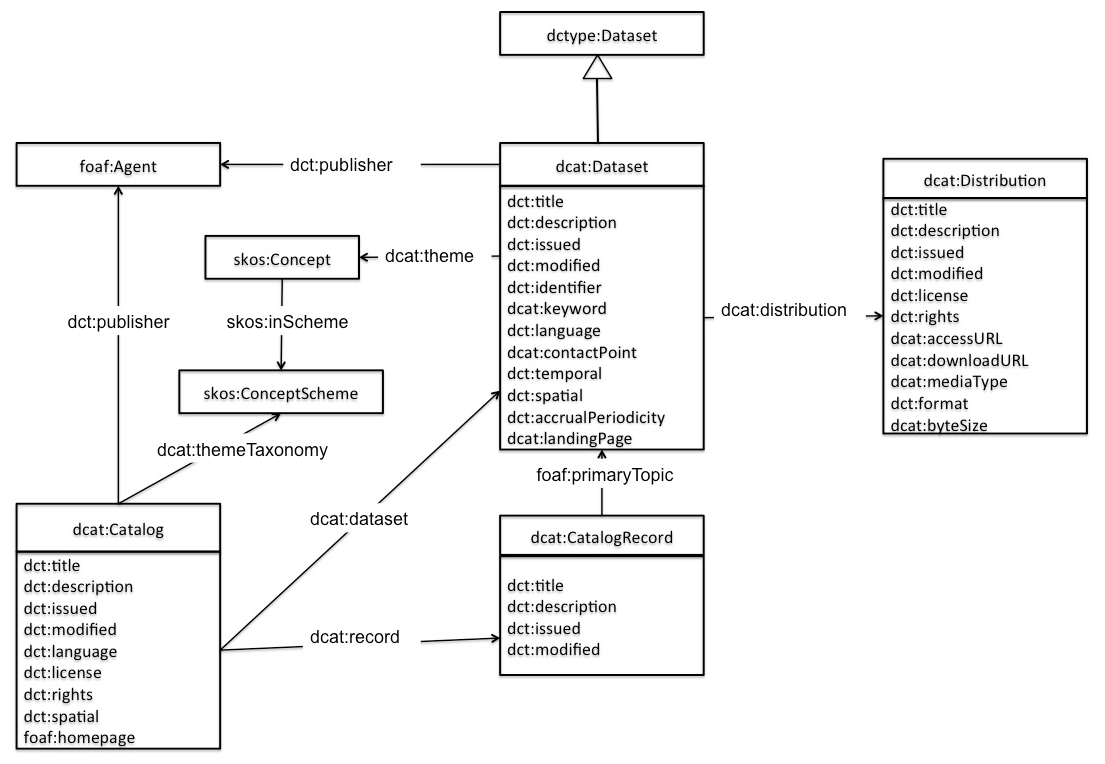


Figure : DCAT schematic data model

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

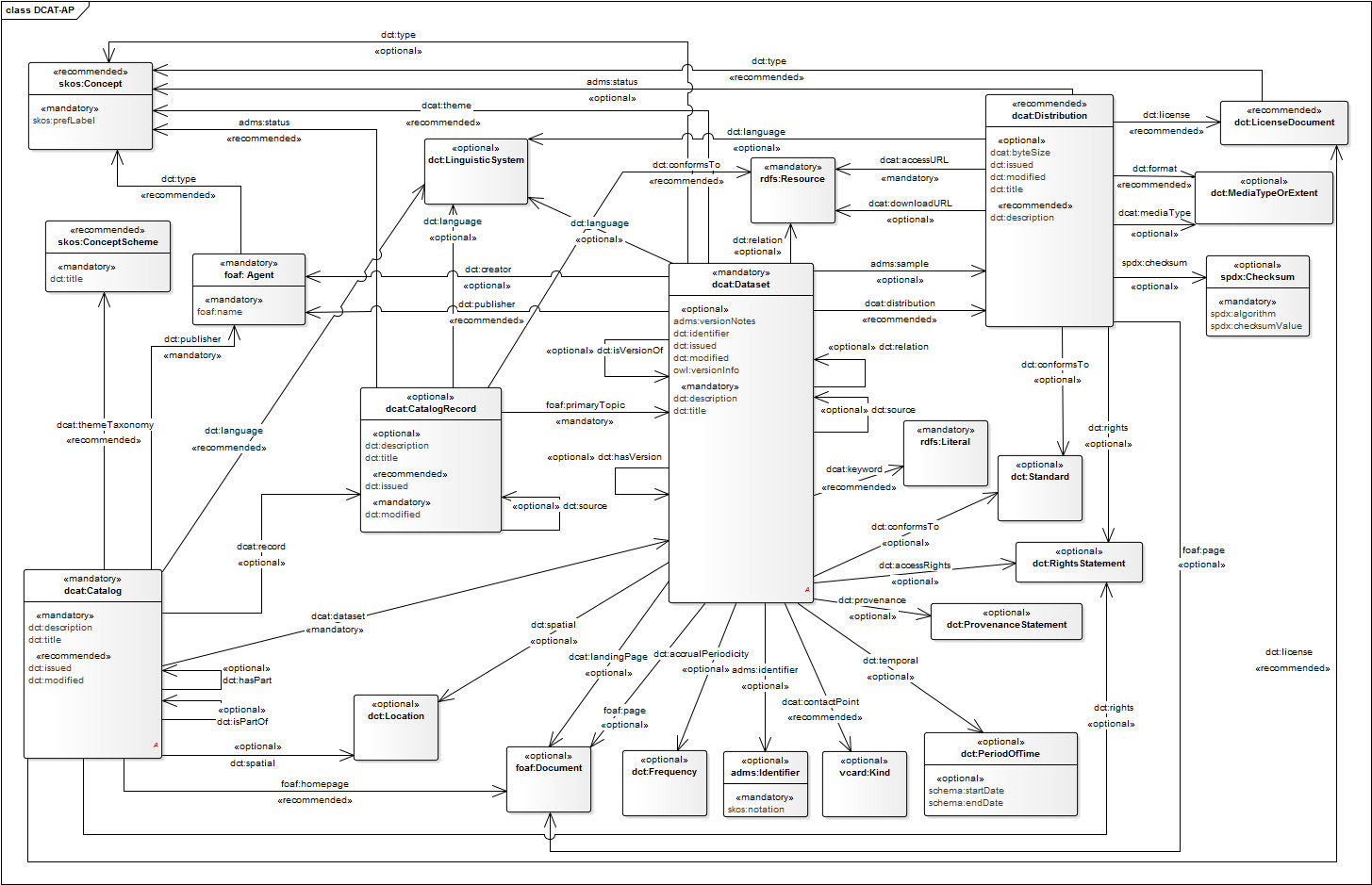


Figure 5: DCAT-AP Data Model

### 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]](#footnote-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]](#footnote-23) of the GeoDCAT-AP was published in December 2015.

### 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.

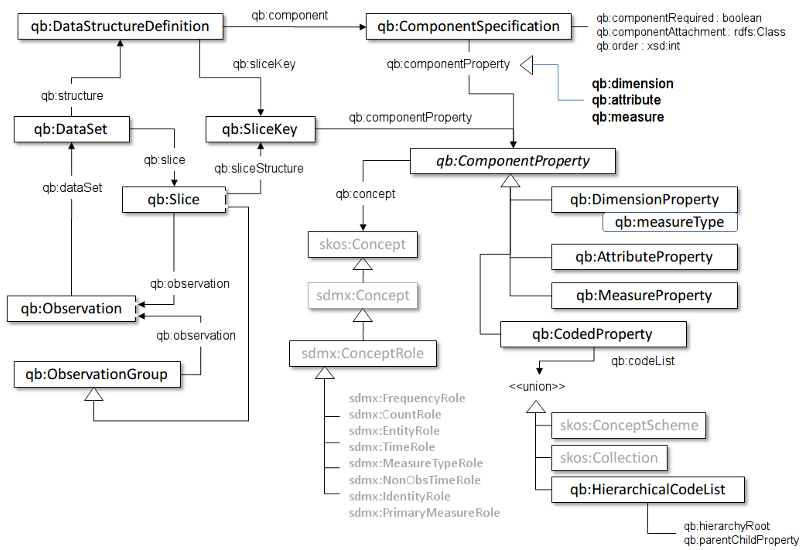


Figure : Data Cube vocabulary overview of key terms and relationships

# Use cases

## 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.

## 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]](#footnote-24) exists. Work on StatDCAT-AP may, however, include discussions at this level since it may improve insight.

# Methodology

## ISA Core Vocabulary process and methodology

This work is conducted according to a process and methodology[[25]](#footnote-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>

## 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

## 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.

## 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 |
| 30 August through 31 October 2016 | public review period |
| 14 November 2016 | fifth virtual WG meeting to discuss and resolve public comments received |
| December 2016 | approval of StatDCAT-AP version 1 for publication |

# The StatDCAT-AP data model

## 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]](#footnote-26), and can be used to describe datasets in any format, for example, those published in SDMX, Data Cube, CSV and other formats.

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]](#footnote-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.

## 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 during the development of StatDCAT-AP can be seen at <https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/issue/all>.

### Dimensions and 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). |

|  |  |
| --- | --- |
| 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. |

Both these properties can be used irrespective of the format of the Distribution. Similar properties have been defined for the Data Cube Vocabulary, namely qb:attribute and qb:dimension. While a proposal was made to use these existing properties, it was decided that using these properties in StatDCAT-AP – as properties of a dcat:Dataset – would not be in line with the way they are defined in the Data Cube Vocabulary – which is in the specific context of a qb:DataStructureDefinition. However, if a Distribution is expressed as a Data Cube DataSet, the values provided for qb:attribute and qb:dimension properties can be used also for the values of stat:attribute and stat:dimension. If a Distribution is expressed in some other format (e.g. SDMX), the values for the properties stat:attribute and stat:dimension can be extracted from corresponding elements in that format if such elements exist. The use of the properties for all Distribution formats ensures coherence across descriptions of Datasets, irrespective of the Distribution format.

|  |
| --- |
| *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#>.* |

### 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]](#footnote-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>.* |

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

### 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]](#footnote-29).

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 | dct: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*](http://nationalportal.org/data/employment)*/emplvis> a dcat:Distribution.*  *dct:type <http://publications.europa.eu/resource/authority/distribution-type/VISUALIZATION> ;*  *dcat:accessURL <*[*http://nationalportal.org/data/employment*](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.

### Number of data series

A requirement was brought forward to allow expression of the number of data series contained in a dataset. A new property has been created for this in the StatDCAT-AP namespace.

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.

|  |  |
| --- | --- |
| Property | Number of data series |
| URI | stat:numSeries |
| Subproperty of | dct:extent |
| Range | dct:SizeOrDuration, expressed as xsd:integer |
| Definition | The actual number of series in the data set as referenced in the Distribution. |
| Comment | The Cartesian Product is the number of modalities of each dimension, excluding what Data Cube calls the measure dimension (that denotes which particular measure is being conveyed by the observation). This is the total of the theoretical number of combinations of dimension values. The actual number of series (as represented in numSeries) might be less than the theoretical number calculated as the Cartesian Product, and when combined with the dimension list, a useful indication of the detail of the data in the data set )often called the ‘density’). |

|  |
| --- |
| *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.* |

### Unit of measurement

A requirement was brought forward to allow expression of the unit of measurement. A new property has been created for this in the StatDCAT-AP namespace.

|  |  |
| --- | --- |
| Property | Unit of measurement |
| URI | stat:statUnitMeasure |
| 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:statUnitMeasure of StatDCAT-AP.  *<http://nationalportal.org/data/employment> a dcat:Dataset;*  *stat:statUnitMeasure <http://example.com/measures/percentage> .* |

### Specifying the length of time series

The StatDCAT Application Profile does not specify short-hand notations for time series, such as the notation used among National Statistical Institutes and Eurostat, e.g. 2016Q2 for a quarterly time series starting in the third quarter of 2006 for which the second quarter of 2016 is the latest quarter covered.

To describe such a time series, a StatDCAT-AP-compliant description would specify:

dct:accrualPeriodicity <http://publications.europa.eu/resource/authority/frequency/QUARTERLY> ;

dct:temporal [ schema:startDate “2006-07-01”^^xsd:date ;

schema:endDate “2016-06-30”^^xsd:date ] .

## 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]](#footnote-30).

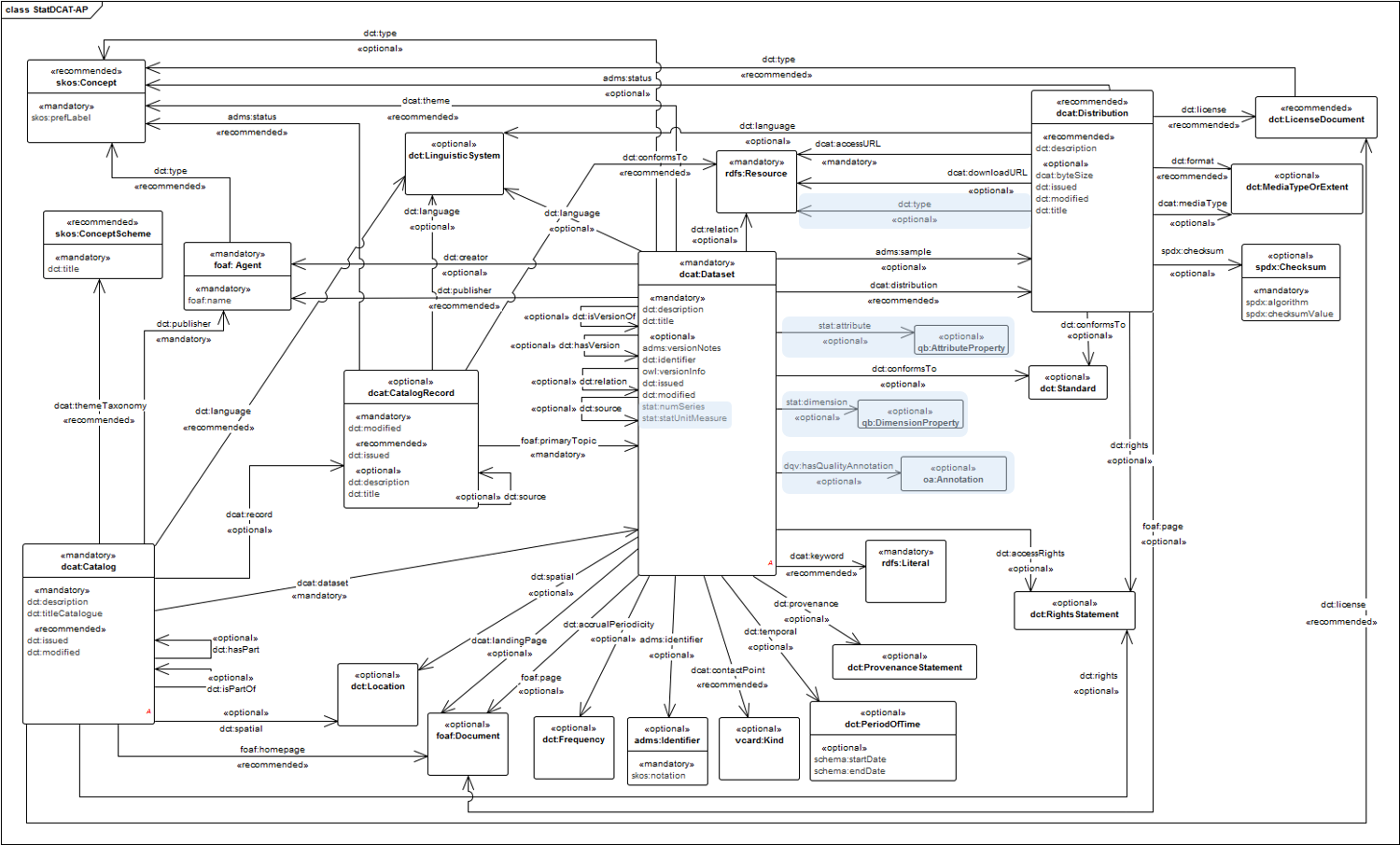
In the given context, the term "processing" means that receivers must accept incoming data and transparently provide these data to applications and services. It does neither imply nor prescribe what applications and services finally do with the data (parse, convert, store, make searchable, display to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

## 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:

|  |  |
| --- | --- |
| Prefix | Namespace URI |
| adms | http://www.w3.org/ns/adms# |
| dcat | http://www.w3.org/ns/dcat# |
| dct | http://purl.org/dc/terms/ |
| dqv | http://www.w3.org/ns/dqv# |
| foaf | http://xmlns.com/foaf/0.1/ |
| oa | http://www.w3.org/ns/oa# |
| qb | http://purl.org/linked-data/cube# |
| rdfs | http://www.w3.org/2000/01/rdf-schema# |
| schema | http://schema.org/ |
| skos | http://www.w3.org/2004/02/skos/core# |
| spdx | http://spdx.org/rdf/terms# |
| stat | http://data.europa.eu/(xyz)/statdcat-ap/[[31]](#footnote-31) |
| vcard | http://www.w3.org/2006/vcard/ns# |
| xsd | http://www.w3.org/2001/XMLSchema# |

## UML Class diagram



## Description of classes

### 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[[32]](#footnote-32) 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[[33]](#footnote-33). | 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> |

### 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.

### Optional Classes

|  |  |  |  |
| --- | --- | --- | --- |
| Class name | Usage note for the Application Profile | URI | Reference |
| Catalogue Record | A description of a Dataset’s entry in the Catalogue. | dcat:CatalogRecord | <http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog-record> |
| Checksum | A value that allows the contents of a file to be authenticated. This class allows the results of a variety of checksum and cryptographic message digest algorithms to be represented. | spdx:Checksum | <http://spdx.org/rdf/terms#Checksum> |
| Document | A textual resource intended for human consumption that contains information, e.g. a web page about a Dataset. | foaf:Document | <http://xmlns.com/foaf/spec/#term_Document> |
| Frequency | A rate at which something recurs, e.g. the publication of a Dataset. | dct:Frequency | <http://dublincore.org/documents/dcmi-terms/#terms-Frequency> |
| Identifier | An identifier in a particular context, consisting of the string that is the identifier; an optional identifier for the identifier scheme; an optional identifier for the version of the identifier scheme; an optional identifier for the agency that manages the identifier scheme | adms:Identifier | <http://www.w3.org/TR/vocab-adms/#identifier> |
| Kind | A description following the vCard specification, e.g. to provide telephone number and e-mail address for a contact point. Note that the class Kind is the parent class for the four explicit types of vCard (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[[34]](#footnote-34) 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> |
| 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> |

## 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.

### 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 11. | 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 |

### 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 |

### 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. | 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[[35]](#footnote-35), DataCite[[36]](#footnote-36), DOI[[37]](#footnote-37), EZID[[38]](#footnote-38) or W3ID[[39]](#footnote-39). | 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 |
| 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). Attribute is a ‘conceptual’ entity that applies to all distribution formats, e.g. in case a dataset is provided both in SDMX and in Data Cube. | 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. Dimension is a ‘conceptual’ entity that applies to all distribution formats, e.g. in case a dataset is provided both in SDMX and in Data Cube. | 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:statUnitMeasure | 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. Unit of measurement is a ‘conceptual’ entity that applies to all distribution formats, e.g. in the case when a dataset is provided both in SDMX and in Data Cube. | 0..n |

### 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 |

### 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 |

### 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 |

### 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 |

### 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 |

### 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[[40]](#footnote-40) | This property contains a string that is an identifier in the context of the identifier scheme referenced by its datatype. | 0..1 |

### 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 |

### 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). | | | | |

## 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.

### 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.

### 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[[41]](#footnote-41) | <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> | 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> | 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:accessRights | Dataset | MDR Access Rights Named Authority List[[42]](#footnote-42) | <http://publications.europa.eu/resource/authority/access-right/> |  |
| dct:accrualPeriodicity | Dataset | MDR Frequency Named Authority List[[43]](#footnote-43) | <http://publications.europa.eu/resource/authority/frequency/> |  |
| dct:format | Distribution | MDR File Type Named Authority List[[44]](#footnote-44) | <http://publications.europa.eu/resource/authority/file-type/> |  |
| dct:language | Catalogue, Dataset | MDR Languages Named Authority List[[45]](#footnote-45) | <http://publications.europa.eu/resource/authority/language/> |  |
| dct:publisher | Catalogue, Dataset | MDR Corporate Bodies Named Authority List[[46]](#footnote-46) | <http://publications.europa.eu/resource/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[[47]](#footnote-47), MDR Countries Named Authority List[[48]](#footnote-48), MDR Places Named Authority List[[49]](#footnote-49), Geonames | <http://publications.europa.eu/resource/authority/country/>, <http://publications.europa.eu/resource/authority/place/>,  <http://publications.europa.eu/resource/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](http://purl.org/adms/status/)/ | :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[[50]](#footnote-50) |
| 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 included in the ADMS specification |

### 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’

### 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[[51]](#footnote-51), the CERIF standard vocabularies[[52]](#footnote-52), the Dewey Decimal Classification[[53]](#footnote-53) and numerous other schemes.

### Licence vocabularies

Concerning licence vocabularies, implementers are encouraged to use widely recognised licences such as Creative Commons licences[[54]](#footnote-54), and in particular the CC Zero Public Domain Dedication[[55]](#footnote-55), the Open Data Commons Public Domain Dedication and License (PDDL)[[56]](#footnote-56), the ISA Open Metadata Licence[[57]](#footnote-57), the European Union Public Licence (EUPL)[[58]](#footnote-58) or an open government licence such as the UK Open Government Licence[[59]](#footnote-59).

Further activities in this area are undertaken by the Open Data Institute[[60]](#footnote-60) with the Open Data Rights Statement Vocabulary[[61]](#footnote-61) and by the Open Digital Rights Language (ODRL) Initiative[[62]](#footnote-62).

# 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.

# Conformance statement

## 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.

## 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.).

# 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, particularly 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[[63]](#footnote-63) maintained by the Publications Office of the EU. Other domain-specific sources for additional properties are the INSPIRE Responsible Party roles[[64]](#footnote-64), the Library of Congress’ MARC relators[[65]](#footnote-65) and DataCite’s contributor types[[66]](#footnote-66). 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[[67]](#footnote-67) 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[[68]](#footnote-68). 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 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.

# Date and Time

Throughout the specification, properties that have values related to temporal aspects are all defined with a range of

**rdfs:Literal typed as xsd:date or xsd:dateTime**

In all these cases, the value can be either of the format “2011-12-05” (xsd:date) or “2011-12-05T13:10:25” (xsd:dateTime).

The choice between these formats in instance metadata depends on the granularity of the information available for the entity that is described. For example, if a dataset is published only once or with a low frequency, only the publication date may be available without the exact time; however, if a dataset is published more than once per day, more precise information about the actual time it was published will be available.

Offering this choice allows the provision of the information as available. Limiting the choice to one of the options would require addition of time information where it is not available – e.g. adding “T00:00:00” or “T23:59:59” – or removal of relevant time information where it is available. Both consequences of limiting the choice, either requiring addition of irrelevant information or deletion of relevant information, are undesirable.

In any case, the information provided must be property typed, e.g.:

dct:modified "2011-12-05"^^**xsd:date**

or:

dct:modified "2011-12-05T13:10:25"^^**xsd:dateTime**

Allowing this flexibility in the provision of information comes at the price of requiring receivers of information to be able to process both formats. This is already a requirement for the basic DCAT-AP, and fully in line with the specification of the W3C Recommendation of the Data Catalog Vocabulary.

# 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](http://www.w3.org/TR/rdf11-concepts/#section-Graph-Literal) are defined by [BCP47](http://tools.ietf.org/html/bcp47)[[69]](#footnote-69), which allows the use of the "t" extension for text transformations defined in [RFC6497](http://tools.ietf.org/html/rfc6497)[[70]](#footnote-70) with the field ["t0"](http://unicode.org/cldr/trac/browser/trunk/common/bcp47/transform_mt.xml)[[71]](#footnote-71) 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[[72]](#footnote-72) 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.

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

# 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. |
| Aurelien Bonnet | IWEPS, Belgium. |
| 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. |

1. 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 |  |  |
| Annotation | oa:Annotation |  |  |  |  |
| Attribute Property | qb:AttributeProperty |  |  |  |  |
| 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 |  |
| Category | skos:Concept | skos:prefLabel |  |  |  |
| Category Scheme | skos:ConceptScheme | 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 | dct:accessRights  dct:conformsTo  foaf:page  dct:accrualPeriodicity  dct:hasVersion  dct:identifier  dct:isVersionOf  dcat:landingPage  dct:language  adms:identifier  dct:provenance  dct:relation  dct:issued  adms:sample  dct:source  dct:spatial  dct:temporal  dct:type  dct:modified  owl:versionInfo  adms:versionNotes | stat:attribute  stat:dimension  stat:numSeries  dqv:hasQualityAnnotation  stat:statUnitMeasure |
| Dimension Property | qb:DimensionProperty |  |  |  |  |
| Distribution | dcat:Distribution | dcat:accessURL | dct:description  dct:format  dct:license | dcat:byteSize  spdx:checksum  foaf:page  dcat:downloadURL  dct:language  dct:conformsTo  dcat:mediaType, subproperty of dct:format  dct:issued  dct:rights  adms:status  dct:title  dct:modified | dct:type |
| Document | foaf:Document |  |  |  |  |
| Frequency | dct:Frequency |  |  |  |  |
| Identifier | adms:Identifier | skos:notation |  |  |  |
| Kind | vcard:Kind |  |  |  |  |
| Licence Document | dct:LicenseDocument | dct:type |  |  |  |
| 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 |  |
| Provenance Statement | dct:ProvenanceStatement |  |  |  |  |
| Publisher Type | skos:Concept |  |  |  |  |
| Resource | rdfs:Resource |  |  |  |  |
| Rights Statement | dct:RightsStatement |  |  |  |  |
| Size or duration | dct:SizeOrDuration |  |  |  |  |
| Standard | dct:Standard |  |  |  |  |
| Status | skos:Concept |  |  |  |  |

1. Mapping SDMX to DCAT
   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.
   1. Diagrams

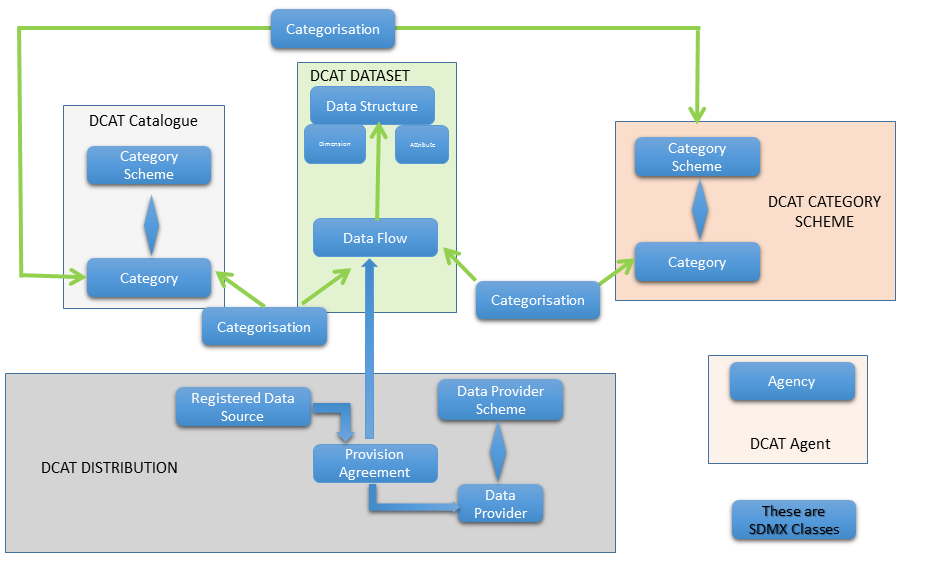


Figure : Schematic map of SDMX Classes to DCAT-AP

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.

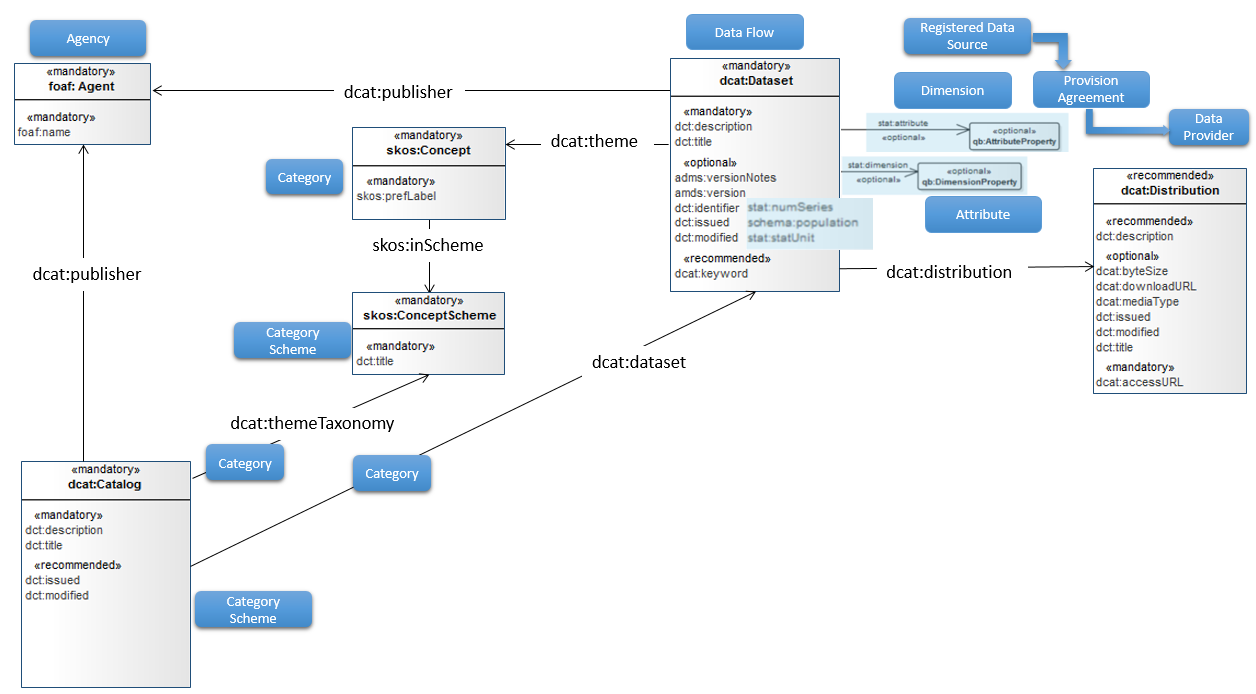


Figure : DCAT-AP Model mapped to SDMX Model Classes

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

* 1. Example
     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.

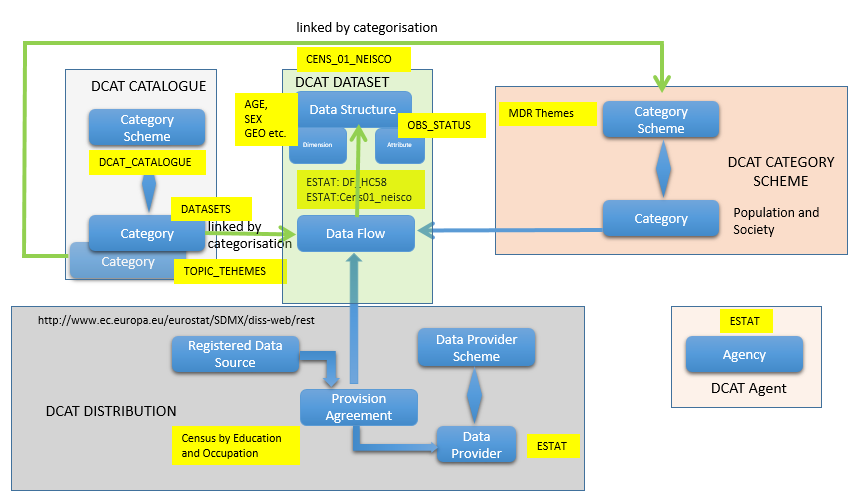


Figure : Metadata Used in the Example Mapping

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*](http://www.ec.europa.eu/eurostat/SDMX/diss-web/rest)) is the dcat:accessURL in the DCAT Distribution.
   * 1. 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:

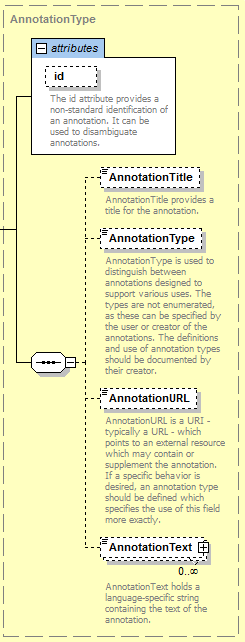


Figure : SDMX XML schema specification for Annotation

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)

* + 1. 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. 

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

e.g. 

* + 1. Data Catalogue

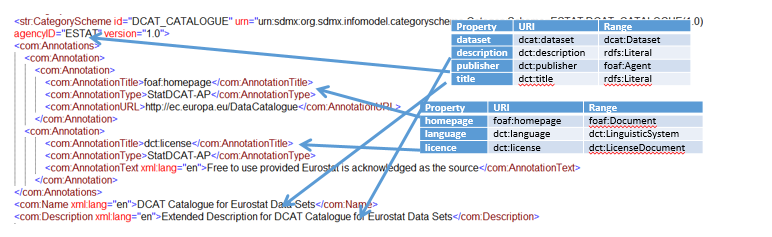
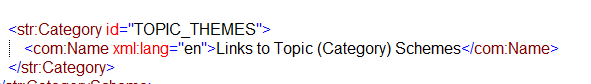
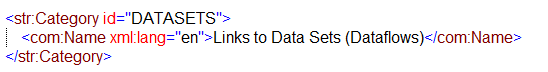
 

Figure : SDMX-DCAT mapping example for the DCAT Catalogue

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

* + 1. Linking to Categories using Categorisations

*Schematic*

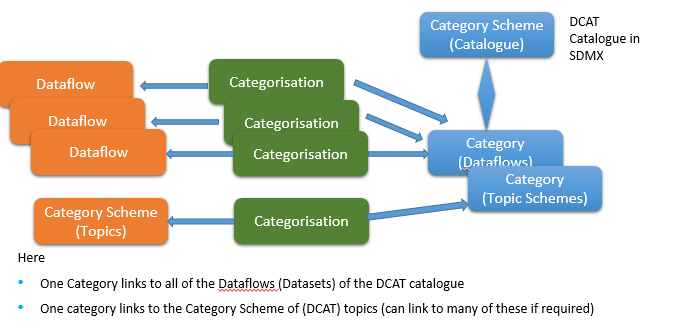


Figure : Schematic showing linking of SDMX Categories to other SDMX objects

*Example*

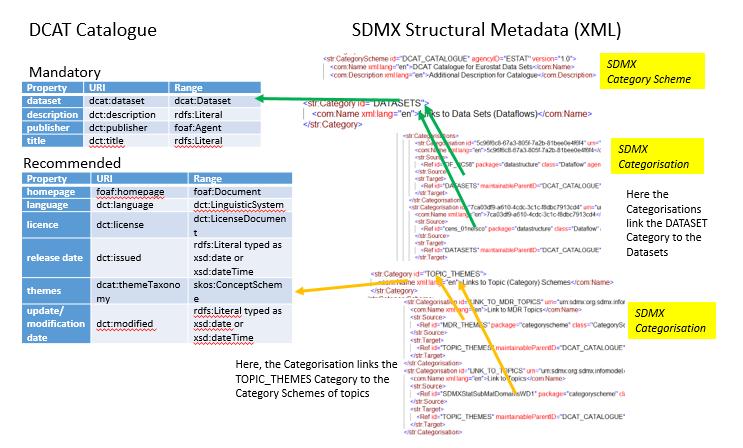


Figure : Linking Catalogue to DCAT Datasets and Category (Topic) Scheme

|  |  |
| --- | --- |
| 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) |

* + 1. Dataset

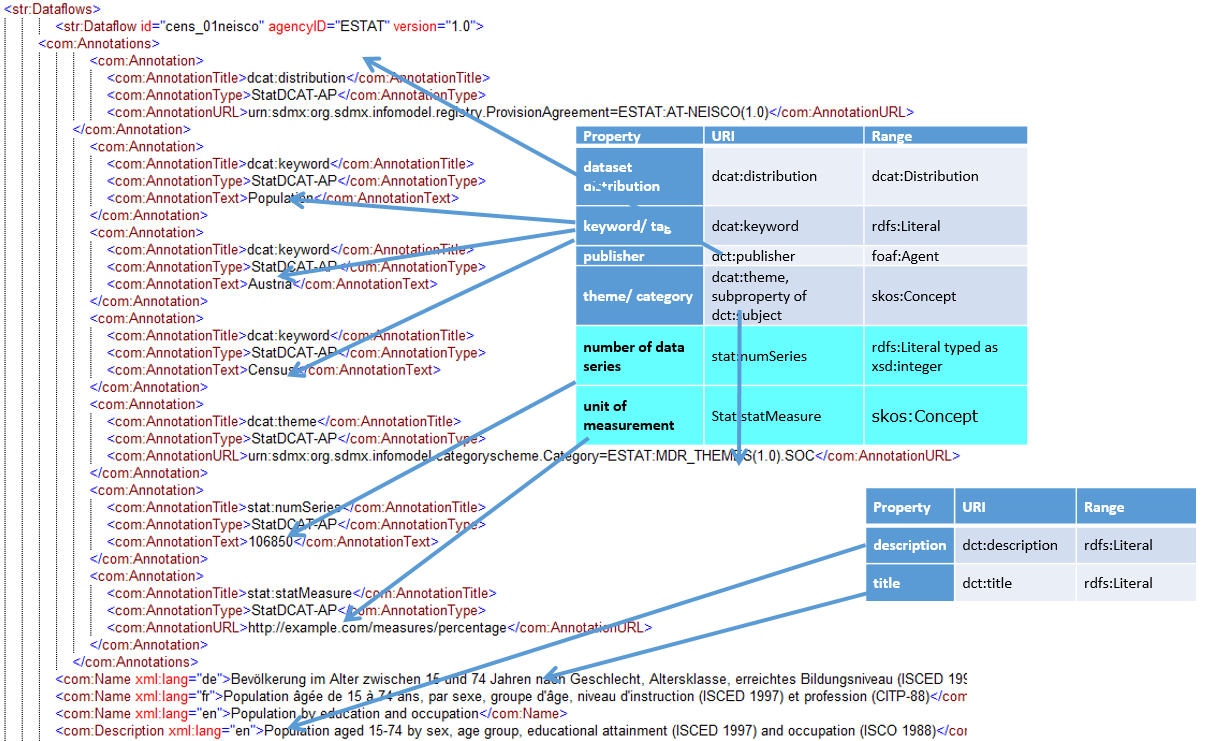


Figure : SDMX to DCAT mapping example for the StatDCAT-AP Dataset

|  |  |
| --- | --- |
| 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:statUnitMeasure | AnnotationURL |
| dct:description | Description |
| dct:title | Name |

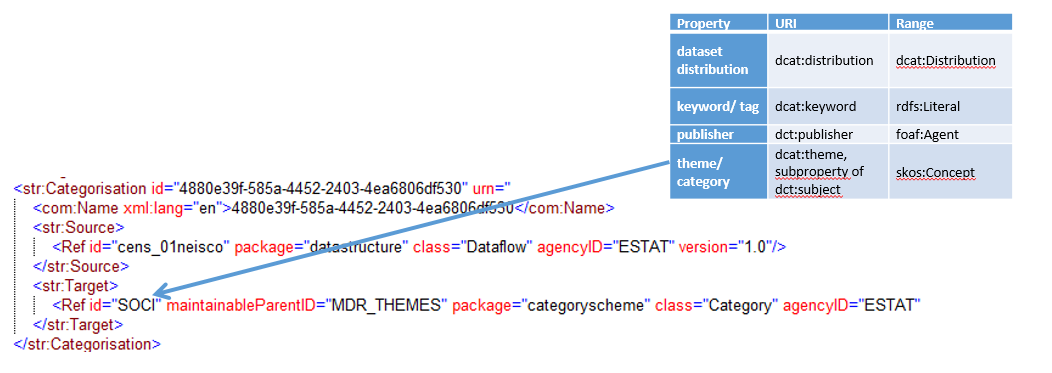


Figure : Linking a Dataflow to the SDMX Category (Topic)

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.

* + 1. Dimension Property and Attribute Property

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

* + 1. Quality Annotation

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

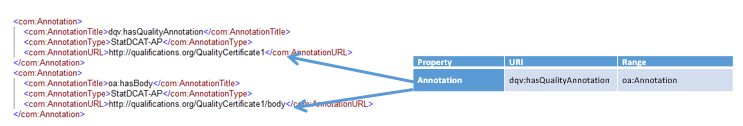


Figure : SDMX to DCAT mapping example for the StatDCAT-AP Annotation

* + 1. Distribution

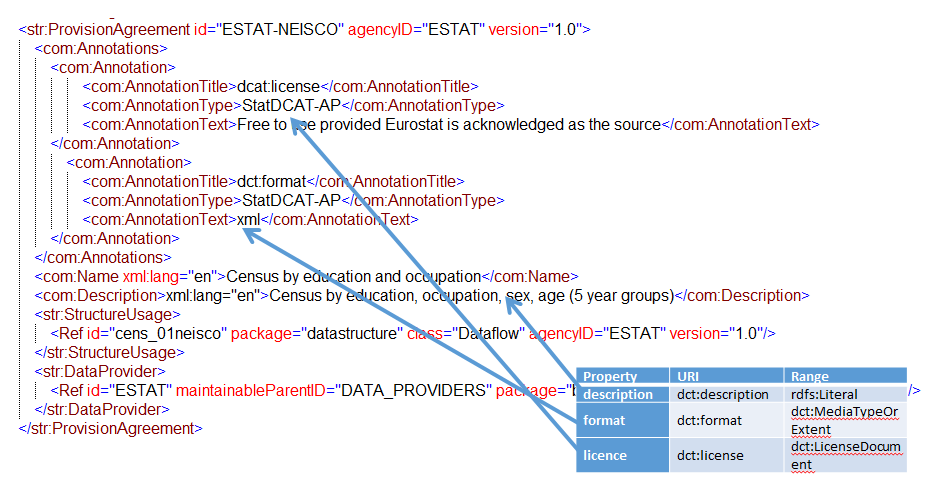


Figure : Linking a Distribution to the SDMX Provision Agreement

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

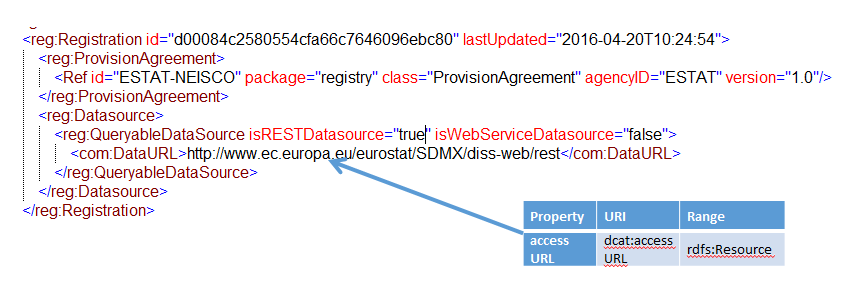


Figure : Linking a Distribution (accessURL) to the SDMX Provision Agreement

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

* + 1. Agent

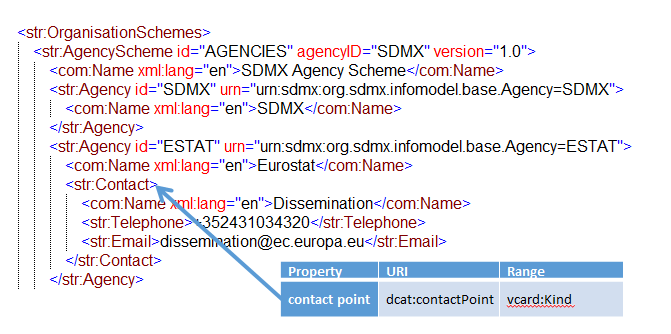


Figure :Linking an Agent to the SDMX Agency

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

* 1. 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).

1. SDMX-based Transformation Mechanism
   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.

* 1. Transformation mechanism

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

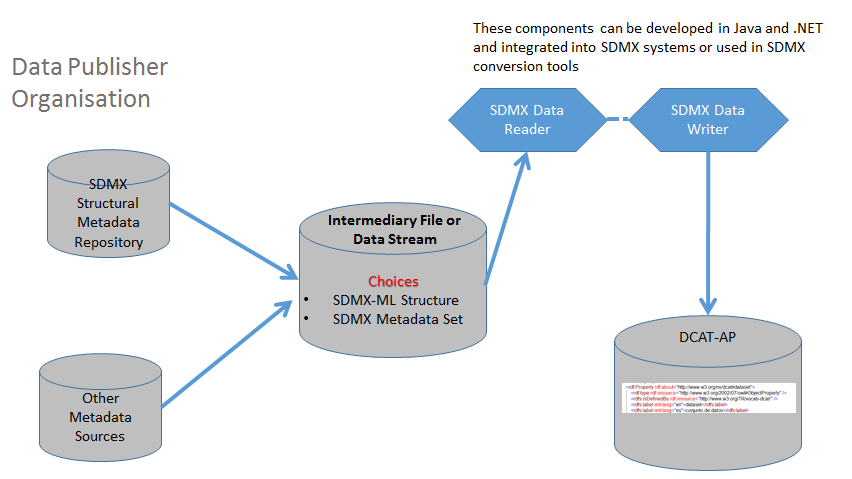


Figure : Diagram of the flow of metadata though the Intermediary Mechanism

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.

* 1. Transformation input formats
     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.

* + 1. 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:

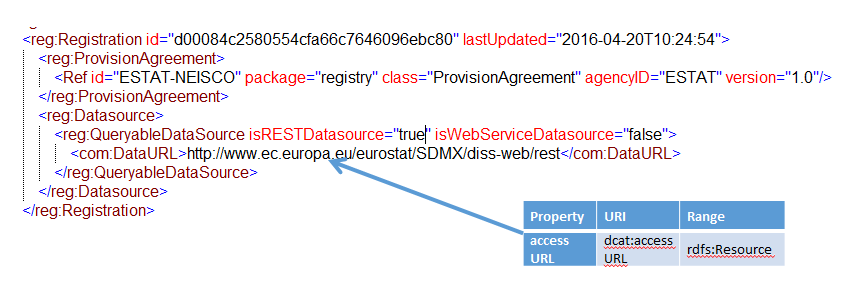


Figure : From Section 9 - Linking a Distribution (accessURL) to the SDMX Provision Agreement

Using the Transformation Mechanism this is output as:



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

So, the full example of the Provision Agreement is:

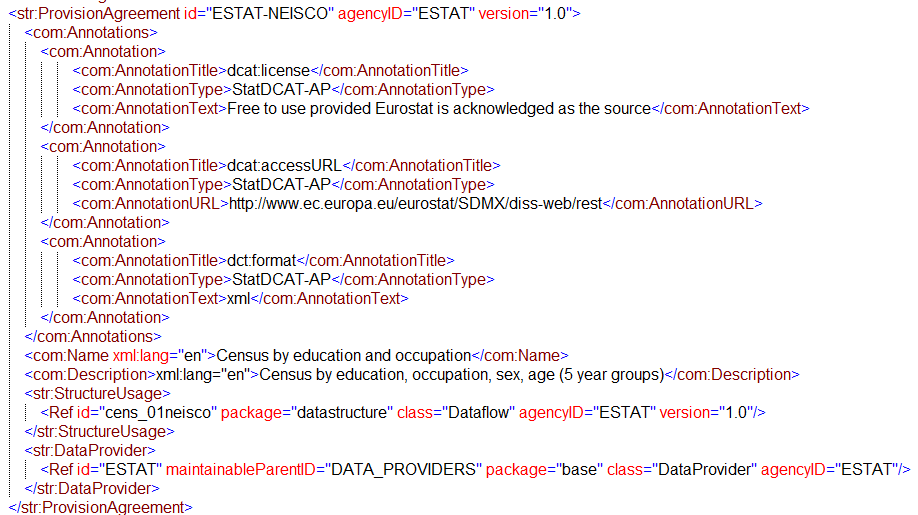


Figure : Example Provision Agreement for DCAT-AP Distribution

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.

* + 1. 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.

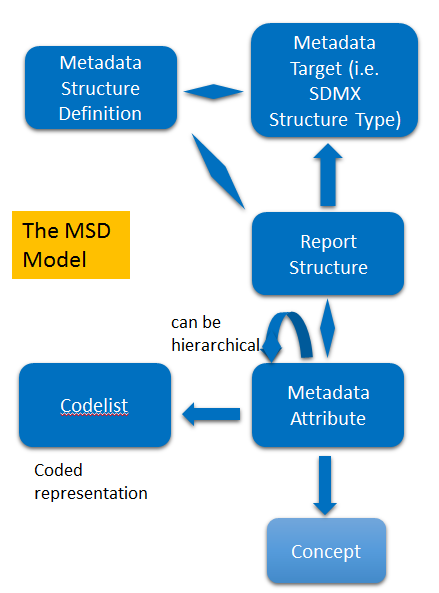


Figure : Schematic diagram of the SDMX Metadata Structure Definition

A schematic of the Metadata Set is shown below.

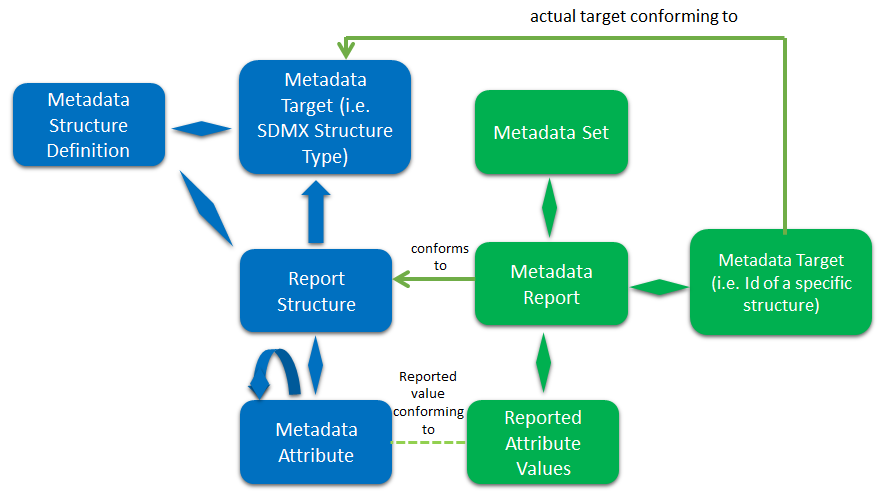


Figure : Schematic diagram of the SDMX Metadata Set

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 DCAT-AP, and the extensions added by StatDCAT-AP.

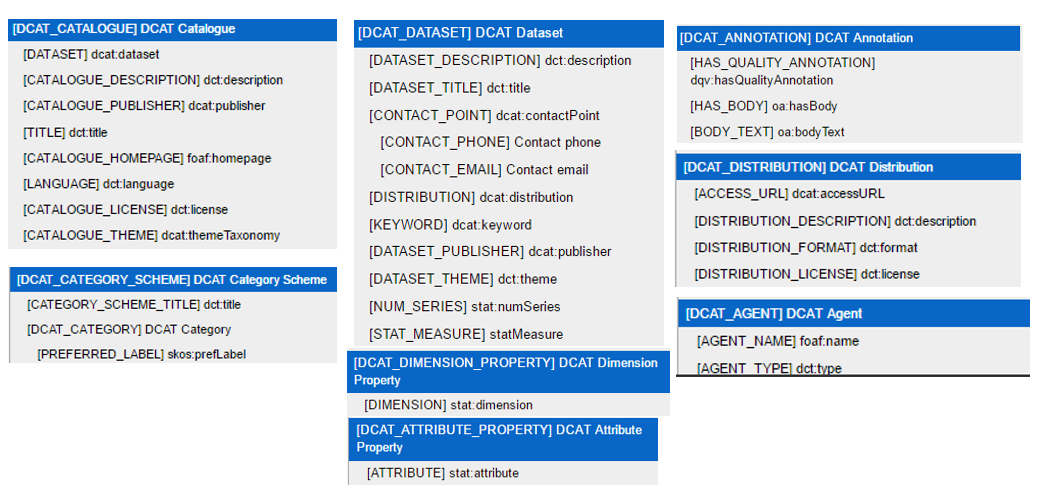


Figure : Metadata Attributes in the DCAT-AP MSD

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.

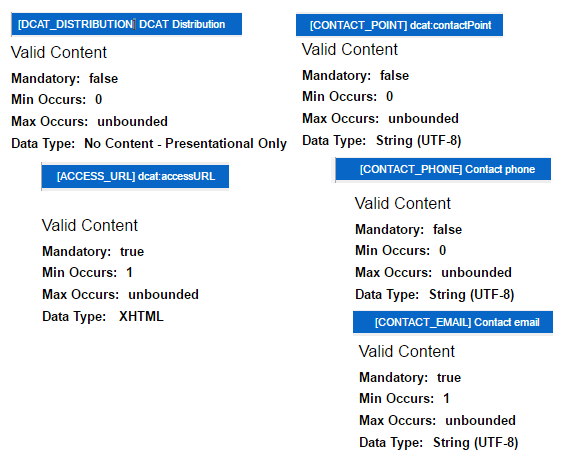


Figure : Example Metadata Attribute Specification

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.

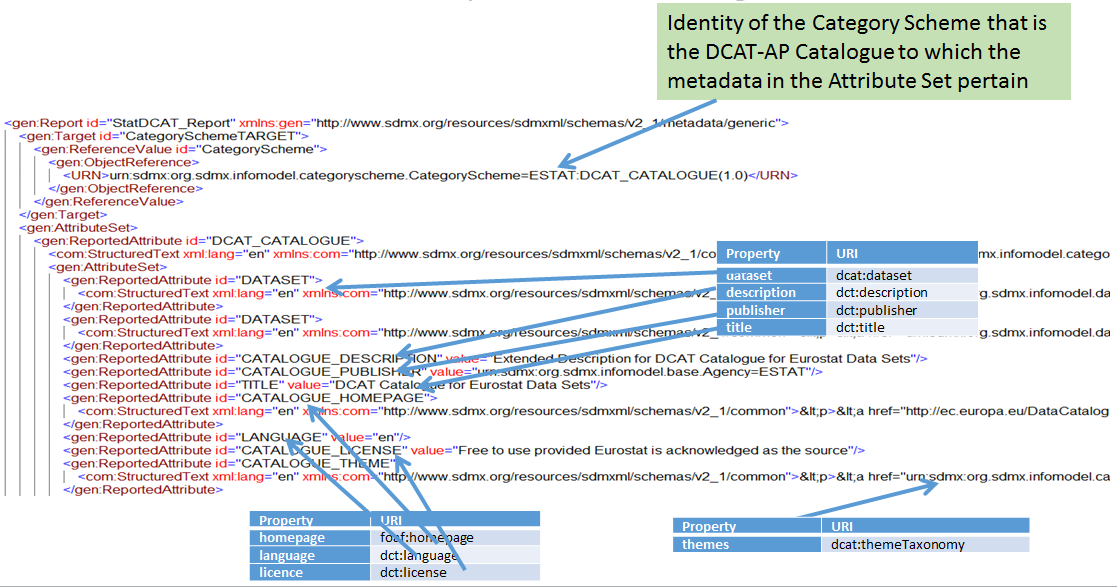


Figure : SDMX catalogue metadata pertaining to the DCAT-AP Catalogue

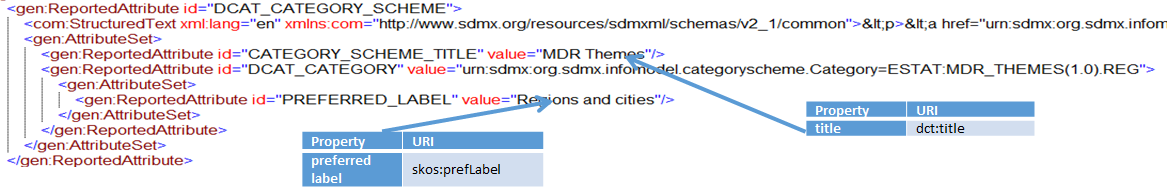


Figure : SDMX category scheme metadata pertaining to the DCAT-AP Catalogue

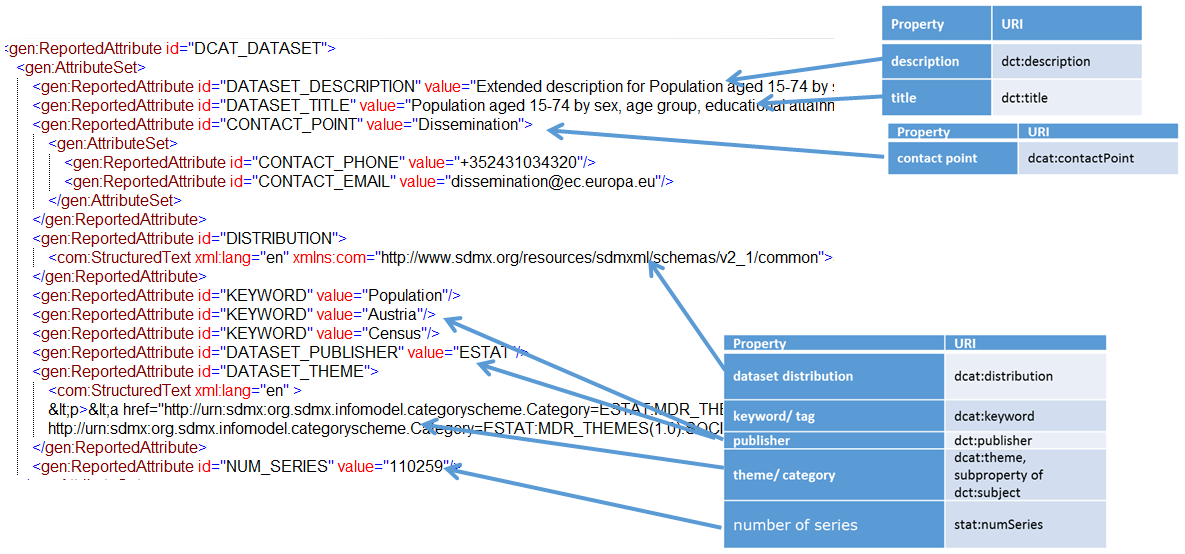


Figure : SDMX dataset metadata pertaining to the DCAT-AP Catalogue including StatDCAT-AP extensions to the Dataset.

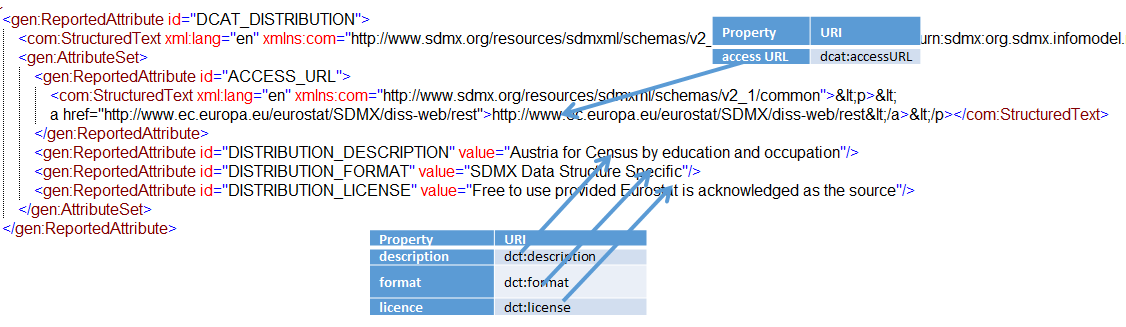


Figure : SDMX distribution metadata pertaining to the DCAT-AP Catalogue including StatDCAT-AP extensions to the Distribution

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.

* 1. Advantages and disadvantages of the two transformation formats
     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.

* + 1. 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
  1. 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.

1. SDMX Files used for the examples
   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:statUnitMeasure</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:AnnotationURL>

</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>

* 1. SDMX Metadata Set
     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>

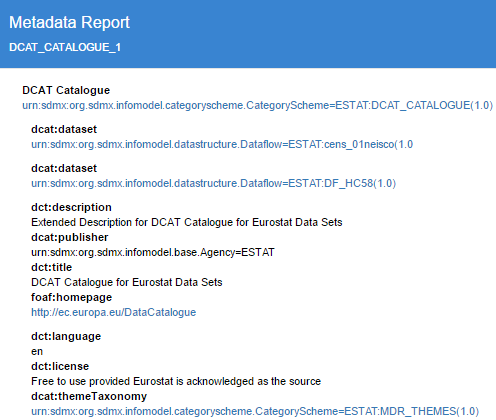
</gen:ObjectReference>

</gen:ReferenceValue>

</gen:Target>

<gen:AttributeSet>

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



<gen:ReportedAttribute id="DCAT\_CATALOGUE">

<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.categoryscheme.CategoryScheme=ESTAT:DCAT\_CATALOGUE(1.0)">urn: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>&lt;a href="urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:cens\_01neisco(1.0">urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:cens\_01neisco(1.0&lt;/a>&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>&lt;a href="urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=ESTAT:DF\_HC58(1.0)">urn:sdmx:org.sdmx.infomodel.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>&lt;a href="http://ec.europa.eu/DataCatalogue">http://ec.europa.eu/DataCatalogue&lt;/a>&lt;/p></com:StructuredText>

</gen:ReportedAttribute>

<gen:ReportedAttribute id="LANGUAGE" value="en"/>

<gen:ReportedAttribute id="CATALOGUE\_LICENSE" value="Free to use provided Eurostat is acknowledged as the source"/>

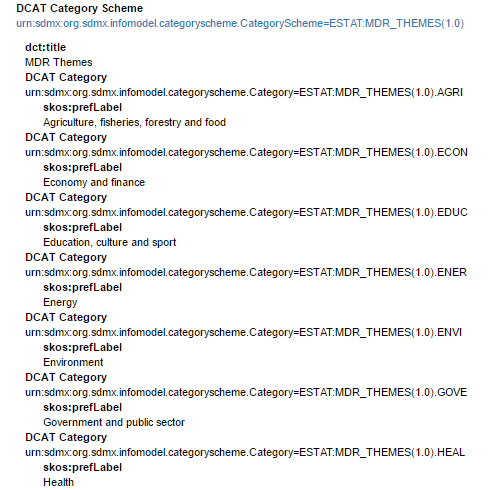
<gen:ReportedAttribute id="CATALOGUE\_THEME">

<com:StructuredText xml:lang="en" xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2\_1/common">&lt;p>&lt;a href="urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR\_THEMES(1.0)">urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR\_THEMES(1.0)&lt;/a>&lt;/p></com:StructuredText>

</gen:ReportedAttribute>

</gen:AttributeSet>

</gen:ReportedAttribute>



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>&lt;a href="urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR\_THEMES(1.0)">urn:sdmx:org.sdmx.infomodel.categoryscheme.CategoryScheme=ESTAT:MDR\_THEMES(1.0)&lt;/a>&lt;/p></com:StructuredText>

<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….



<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">&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: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">&lt;p>&lt;a href="http://urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR\_THEMES(1.0).SOCI">http://urn:sdmx:org.sdmx.infomodel.categoryscheme.Category=ESTAT:MDR\_THEMES(1.0).SOCI&lt;/a>&lt;/p></com:StructuredText>

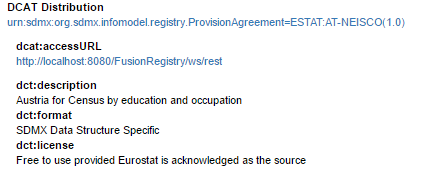
</gen:ReportedAttribute>

<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">&lt;p>&lt;a href="http://example.com/measures/percentage">http://example.com/measures/percentage&lt;/a>&lt;/p></com:StructuredText>

</gen:ReportedAttribute>



<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>

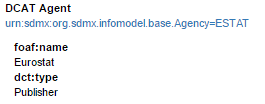
<gen:ReportedAttribute id="DISTRIBUTION\_DESCRIPTION" value="Austria for Census by education and occupation"/>

<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>



<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=ESTAT&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>

1. Examples of StatDCAT-AP descriptions of Data Cube DataSets
   1. RDF Example 1

Derived from <https://www.w3.org/TR/vocab-data-cube/#full-example>

@prefix dcat: <http://www.w3.org/ns/dcat#> .

@prefix dct: <http://purl.org/dc/terms/> .

@prefix eg: <http://example.org/ns#> .

@prefix geonames: <http://sws.geonames.org/> .

@prefix sdmx-dimension: <http://purl.org/linked-data/sdmx/2009/dimension#> .

@prefix sdmx-attribute: <http://purl.org/linked-data/sdmx/2009/attribute#> .

@prefix stat : <http://data.europa.eu/(xyz)/statdcat-ap/> .

@prefix theme: <http://publications.europa.eu/resource/authority/data-theme/> .

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

# -- Dataset --------------------------------------------

eg:dataset-le3 a dcat:Dataset ;

dct:title "Life expectancy"@en ;

dct:description "Life expectancy within Welsh Unitary authorities - extracted from Stats Wales"@en ;

dct:publisher eg:organization ;

dct:issued "2010-08-11"^^xsd:date ;

# These four properties have been copied from the source metadata.

dcat:theme

theme:HEAL , # Health

theme:REGI ; # Regions and cities

# dct:subject

# sdmx-subject:3.2 , # regional and small area statistics

# sdmx-subject:1.4 , # Health

#

# These subjects have been mapped to the MDR Data Theme NAL. Note that this is not a

# trivial mapping but may require specific mapping tables or owl:sameAs assertions.

dct:spatial geonames:2634865 ; #Wales

# ex-geo:wales; # Wales

# The subject in the source mapped to dct:spatial and value from Geonames. Note again that

# this may require specific mapping tables or owl:sameAs assertions.

stat:dimension eg:refArea ;

stat:dimension eg:refPeriod ;

stat:dimension sdmx-dimension:sex ;

stat:attribute sdmx-attribute:unitMeasure ;

# The values of these properties copied from the qb:attribute and qb:dimension properties

# in the source metadata.

stat:statUnitMeasure <http://dbpedia.org/resource/Year> .

# The value of this property copied from the sdmx-attribute:unitMeasure property

# in the source metadata.

# -- Distribution -----------------------------------------

eg:distribution-le3 a dcat:Distribution ;

dcat:accessURL <http://example.org/file.rdf> ;

dct:format <http://publications.europa.eu/mdr/authority/file-type/RDF\_TURTLE> .

# The format is not specified in the source metadata but added here for completeness.

# The URI for Distribution is declared in the eg: namespace – this was not specified in the

# the source metadata as this did not describe the distribution separately.

# In addition, a link to a file is shown as value for dcat:accessURL which was not

# specified in the source metadata.

# Note that the definitions of eg:organization, eg:refArea and eg:refPeriod have not been

# reproduced here. They are defined in the source metadata.

* 1. RDF Example 2

Derived from <https://digital-agenda-data.eu/datasets/digital_agenda_scoreboard_key_indicators/@@get_dataset_metadata>

@prefix adms: <http://www.w3.org/ns/adms#> .

@prefix dad-prop: <http://semantic.digital-agenda-data.eu/def/property/> .

@prefix dcat: <http://www.w3.org/ns/dcat#> .

@prefix dcterms: <http://purl.org/dc/terms/> .

@prefix ns22: <http://semantic.digital-agenda-data.eu/dataset/> .

@prefix ns25: <http://ec.europa.eu/geninfo/legal\_notices\_en.htm#> .

@prefix ns26: <http://publications.europa.eu/resource/authority/access-right/> .

@prefix ns27: <http://publications.europa.eu/resource/authority/language/> .

@prefix ns28: <http://publications.europa.eu/resource/authority/corporate-body/> .

@prefix ns29: <http://publications.europa.eu/resource/authority/country/> .

@prefix ns30: <http://semantic.digital-agenda-data.eu/dataset/digital-agenda-scoreboard-key-indicators/distribution/> .

@prefix ns31: <http://publications.europa.eu/resource/authority/data-theme/> .

@prefix ns33: <http://purl.org/adms/status/> .

@prefix ns35: <http://publications.europa.eu/resource/authority/frequency/> .

@prefix ns38: <http://publications.europa.eu/resource/authority/file-type/> .

@prefix ns39: <http://publications.europa.eu/resource/authority/distribution-type/> .

@prefix ns40: <http://digital-agenda-data.eu/datasets/digital\_agenda\_scoreboard\_key\_indicators#> .

@prefix ns41: <http://digital-agenda-data.eu/datasets/> .

@prefix stat : <http://data.europa.eu/(xyz)/statdcat-ap/> .

@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

# ---- Dataset ----------------------------------------------------

ns22:digital-agenda-scoreboard-key-indicators a dcat:Dataset ;

dcterms:title "Digital Agenda Key Indicators" ;

dcterms:description "European Commission services selected more than 100 indicators, divided into thematic groups, which illustrate some key dimensions of the European information society (Telecom sector, Broadband, Mobile, Internet usage, Internet services, eGovernment, eCommerce, eBusiness, ICT Skills, Research and Development).\nThese indicators allow a comparison of progress across European countries as well as over time.\nYou can also browse the data with the help of a visualisation tools going at http://digital-agenda-data.eu/, where you are also able to download selected information." ;

dcterms:identifier "digital-agenda-scoreboard-key-indicators" ;

dcterms:issued "2011-05-01T00:00:00Z"^^xsd:dateTime ;

dcterms:modified "2016-11-11T10:31:56+02:00"^^xsd:dateTime ;

dcterms:publisher ns28:CNECT ;

dcterms:language ns27:ENG ;

dcterms:accessRights ns26:PUBLIC ;

dcterms:spatial ns29:TUR , ns29:NOR , ns29:ISL , ns29:EUR , ns29:CHE ;

dcat:keyword "ebusiness" , "broadband" , "internet" , "ICT research" , "ecommerce" , "digital agenda" , "telecom market" , "ICT skills" , "information-society" ;

dcat:theme ns31:SOCI , ns31:TECH , ns31:GOVE ;

dcterms:temporal [ schema:startDate "2002-01-01+03:00"^^xsd:date ] ;

dcat:distribution ns30:download ;

dcat:distribution ns30:visualisation ;

adms:status ns33:Completed ;

dcat:landingPage <http://digital-agenda-data.eu/> ;

# all properties above are copies of the source metadata.

# <https://digital-agenda-data.eu/datasets/digital_agenda_scoreboard_key_indicators>.

# However, the duplicated values of dcterms:title and dcterms:description in rdfs:label and

# rdfs:comment have not been included here.

dcat:contactPoint [ vcard:fn "DG CONNECT - Digital Economy and Skills (Unit F.4)" ;

vcard:hasEmail <mailto:CNECT-F4@ec.europa.eu> ;

vcard:hasURL <https://ec.europa.eu/digital-single-market/en/> ] ;

# The information for contact point has been adapted to follow DCAT-AP guidelines

# <https://joinup.ec.europa.eu/node/150343/>, using vcard properties. The source treats contact

# point as both a vcard:Organization and an org:Organization which seems to be sematically

# questionable.

dcterms:accrualPeriodicity ns35:ANNUAL\_2 ;

# The source metadata has the statement

# dcat:accrualPeriodicity ns35:ANNUAL\_2 ; which is incorrect

stat:attribute dad-prop:flag ;

stat:attribute dad-prop:note ;

stat:dimension dad-prop:breakdown ;

stat:dimension dad-prop:indicator ;

stat:dimension dad-prop:ref-area ;

stat:dimension dad-prop:time-period ;

stat:dimension dad-prop:unit-measure .

# The values of the properties stat:attribute and stat:dimensions have been copied from all

# occurrences of qb:attribute and qb:dimension properties as suggested in the StatDCAT-AP

# specification.

# ---- Distributions ----------------------------------------------

ns30:download a dcat:Distribution ;

dcterms:format ns38:RDF\_TURTLE ;

dcterms:type ns39:DOWNLOADABLE\_FILE ;

dcat:accessURL ns40:download ;

dcterms:license ns25:copyright .

ns30:visualisation a dcat:Distribution ;

dcterms:type ns39:VISUALIZATION ;

dcat:accessURL ns41:digital\_agenda\_scoreboard\_key\_indicators ;

dcterms:license ns25:copyright .

# All properties for the distributions have been copied from the source. The only exception

# is dcterms:license. The source assigns this to the dataset; here it is copied to each of

# the distributions in line with DCAT.

# Note that the definitions of the terms in the dad-prop namespace have not been

# reproduced here. They are defined in the source metadata.

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> [↑](#footnote-ref-1)
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> [↑](#footnote-ref-2)
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)> [↑](#footnote-ref-3)
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> [↑](#footnote-ref-4)
5. European Commission. European Data Portal. <http://www.europeandataportal.eu/> [↑](#footnote-ref-5)
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> [↑](#footnote-ref-6)
7. European Commission. ISA – Interoperability Solutions for European Public Administrations. <http://ec.europa.eu/isa/about-isa/> [↑](#footnote-ref-7)
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. [↑](#footnote-ref-8)
9. W3C. The RDF Data Cube Vocabulary. W3C Recommendation 16 January 2014. <https://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/> [↑](#footnote-ref-9)
10. W3C. Data Catalog Vocabulary (DCAT). W3C Recommendation 16 January 2014. <http://www.w3.org/TR/vocab-dcat/> [↑](#footnote-ref-10)
11. Publications Office of the European Union. Metadata Registry (MDR). <http://publications.europa.eu/mdr/> [↑](#footnote-ref-11)
12. Statistical Data and Metadata eXchange. <https://sdmx.org/> [↑](#footnote-ref-12)
13. IETF. RFC 3986. Uniform Resource Identifier (URI): Generic Syntax. <https://www.ietf.org/rfc/rfc3986.txt> [↑](#footnote-ref-13)
14. IETF. RFC 1738. Uniform Resource Locators (URL). <https://www.ietf.org/rfc/rfc1738.txt> [↑](#footnote-ref-14)
15. European Union Open Data Portal. <http://open-data.europa.eu> [↑](#footnote-ref-15)
16. See the file ESTAT\_xxx.zip in <http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing> [↑](#footnote-ref-16)
17. Eurostat. Euro-SDMX Metadata Structure (ESMS). <http://ec.europa.eu/eurostat/data/metadata> [↑](#footnote-ref-17)
18. W3C. Government Linked Data Working Group. <https://www.w3.org/2011/gld/charter> [↑](#footnote-ref-18)
19. US spelling from the original. [↑](#footnote-ref-19)
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> [↑](#footnote-ref-20)
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> [↑](#footnote-ref-21)
22. <http://publications.europa.eu/mdr/eurovoc/> [↑](#footnote-ref-22)
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> [↑](#footnote-ref-23)
24. W3C. The RDF Data Cube Vocabulary. W3C Recommendation 16 January 2014 <http://www.w3.org/TR/vocab-data-cube/> [↑](#footnote-ref-24)
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> [↑](#footnote-ref-25)
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> [↑](#footnote-ref-26)
27. W3C. Data Catalogue Vocabulary. W3C Recommendation 16 January 2014. <https://www.w3.org/TR/vocab-dcat/> [↑](#footnote-ref-27)
28. Data on the Web Best Practices: Data Quality Vocabulary. W3C Working Draft 19 May 2016. <https://www.w3.org/TR/vocab-dqv/> [↑](#footnote-ref-28)
29. MDR Authorities. Distribution types. <http://publications.europa.eu/mdr/authority/distribution-type/> [↑](#footnote-ref-29)
30. IETF. RFC 2119. Key words for use in RFCs to Indicate Requirement Levels. <http://www.ietf.org/rfc/rfc2119.txt> [↑](#footnote-ref-30)
31. The actual namespace URI will be assigned by the European Commission’s URI Committee. [↑](#footnote-ref-31)
32. W3C. The Organization Ontology. W3C Candidate Recommendation, 25 June 2013. <http://www.w3.org/TR/2013/CR-vocab-org-20130625/> [↑](#footnote-ref-32)
33. IETF. BCP 47. Tags for Identifying Languages. <http://www.rfc-editor.org/rfc/bcp/bcp47.txt> [↑](#footnote-ref-33)
34. European Commission. Joinup. Core Location Vocabulary. <https://joinup.ec.europa.eu/asset/core_location/description> [↑](#footnote-ref-34)
35. Mikulski Archive for Space Telescopes (MAST). Referencing Data Sets in Astronomical Literature. <http://archive.stsci.edu/pub_dsn.html> [↑](#footnote-ref-35)
36. DataCite. <http://www.datacite.org/> [↑](#footnote-ref-36)
37. DOI. Digital Object Identifier. <http://www.doi.org/> [↑](#footnote-ref-37)
38. EZID. <http://n2t.net/ezid> [↑](#footnote-ref-38)
39. W3C Permanent Identifier Community Group. Permanent Identifiers for the Web. <https://w3id.org/> [↑](#footnote-ref-39)
40. DataCite Resource Identifier Scheme. <http://purl.org/spar/datacite/ResourceIdentifierScheme> [↑](#footnote-ref-40)
41. Internet Assigned Numbers Authority (IANA). Media Types <http://www.iana.org/assignments/media-types/> [↑](#footnote-ref-41)
42. Publications Office of the European Union. Metadata Registry. Authorities. Access rights. <http://publications.europa.eu/mdr/authority/access-right/> [↑](#footnote-ref-42)
43. Publications Office of the European Union. Metadata Registry. Authorities. Frequencies. <http://publications.europa.eu/mdr/authority/frequency> [↑](#footnote-ref-43)
44. Publications Office of the European Union. Metadata Registry. Authorities. File types. <http://publications.europa.eu/mdr/authority/file-type/> [↑](#footnote-ref-44)
45. Publications Office of the European Union. Metadata Registry. Authorities. Languages. <http://publications.europa.eu/mdr/authority/language/> [↑](#footnote-ref-45)
46. Publications Office of the European Union. Metadata Registry. Authorities. Corporate bodies. <http://publications.europa.eu/mdr/authority/corporate-body/> [↑](#footnote-ref-46)
47. Publications Office of the European Union. Metadata Registry. Authorities. Continents.

    <http://publications.europa.eu/mdr/authority/continent/> [↑](#footnote-ref-47)
48. Publications Office of the European Union. Metadata Registry. Authorities. Countries. <http://publications.europa.eu/mdr/authority/country/> [↑](#footnote-ref-48)
49. Publications Office of the European Union. Metadata Registry. Authorities. Places. <http://publications.europa.eu/mdr/authority/place/> [↑](#footnote-ref-49)
50. European Commission. Joinup. Asset Description Metadata Schema (ADMS). ADMS 1.00. <http://joinup.ec.europa.eu/asset/adms/release/100> [↑](#footnote-ref-50)
51. EuroVoc. <http://eurovoc.europa.eu/>. [↑](#footnote-ref-51)
52. http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.5/CERIF1.5\_Semantics.xhtml [↑](#footnote-ref-52)
53. OCLC. Dewey Summaries as Linked Data. <http://www.oclc.org/dewey/webservices.en.html> and <http://dewey.info/> [↑](#footnote-ref-53)
54. Creative Commons. About The Licenses. <http://creativecommons.org/licenses/> [↑](#footnote-ref-54)
55. Creative Commons. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. <http://creativecommons.org/publicdomain/zero/1.0/> [↑](#footnote-ref-55)
56. Open Data Commons Public Domain Dedication and License (PDDL). <http://opendatacommons.org/licenses/pddl/> [↑](#footnote-ref-56)
57. ISA Open Metadata Licence v1.1, <https://joinup.ec.europa.eu/category/licence/isa-open-metadata-licence-v11> [↑](#footnote-ref-57)
58. European Commission. Joinup. Open Source Software. European Union Public Licence (EUPL). <http://joinup.ec.europa.eu/software/page/eupl> [↑](#footnote-ref-58)
59. The National Archives. Open Government Licence for public sector information. <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/> [↑](#footnote-ref-59)
60. Open Data Institute. <http://www.theodi.org/> [↑](#footnote-ref-60)
61. Open Data Institute. Open Data Rights Statement Vocabulary. <http://schema.theodi.org/odrs/> [↑](#footnote-ref-61)
62. Open Digital Rights Language (ODRL) Initiative. <http://www.w3.org/community/odrl/> [↑](#footnote-ref-62)
63. Publications Office of the European Union. Metadata Registry. Authorities. Roles. <http://publications.europa.eu/mdr/authority/role/> [↑](#footnote-ref-63)
64. European Commission. INSPIRE Registry. Responsible Party Role. <http://inspire.ec.europa.eu/metadata-codelist/ResponsiblePartyRole/> [↑](#footnote-ref-64)
65. Library of Congress. MARC Code List for Relators. <http://loc.gov/marc/relators/relaterm.html> [↑](#footnote-ref-65)
66. 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. [↑](#footnote-ref-66)
67. W3C. PROV-O: The PROV Ontology. W3C Recommendation 30 April 2013. <http://www.w3.org/TR/prov-o/> [↑](#footnote-ref-67)
68. 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> [↑](#footnote-ref-68)
69. Internet Engineering Task Force (IETF). BCP47. Tags for Identifying Languages. <http://tools.ietf.org/html/bcp47> [↑](#footnote-ref-69)
70. Internet Engineering Task Force (IETF). BCP47 Extension T – Transformed Content. <http://tools.ietf.org/html/rfc6497> [↑](#footnote-ref-70)
71. UNICODE Consortium. CLDR - Unicode Common Locale Data Repository. BCP47, transform\_mt.xml. <http://unicode.org/cldr/trac/browser/trunk/common/bcp47/transform_mt.xml> [↑](#footnote-ref-71)
72. Apache Web Server: content negotiation. <http://httpd.apache.org/docs/current/content-negotiation.html> [↑](#footnote-ref-72)