



DCAT Application Profile for data portals in Europe
Draft for Public Review

Date: 13/05/2013



Document Metadata

Property	Value
Date	2013-05-10
Status	Draft Final
Version	0.06
Rights	© 2013 European Union
License	ISA Open Metadata License v1.1, retrievable from https://joinup.ec.europa.eu/category/licence/isa-open-metadata-licence-v11 .
Access URL	This draft specification can be downloaded from the ISA website: https://joinup.ec.europa.eu/asset/dcat_application_profile/description

Disclaimer

This specification was prepared for the ISA programme by:
PwC EU Services

The views expressed in this draft specification 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.

Document History

Version	Date	Description	Action
0.01	2013-05-08	<p>Creation; based on draft 3 with inclusion of proposed resolutions of the issues raised by the Working Group</p> <ul style="list-style-type: none"> • Added text in section 1.2 to explain that the AP is for exchange of data not to require systems to be fully Linked Data compliant. • Added mandatory class LicenseDocument in section 6.1 and section 7.8 • Added ADMS licence type vocabulary as controlled vocabulary for licences in section 8.2 • Added optional 'other identifier' property for Dataset in section 7.3.3 • Added section 11.3 on Provenance information 	Team review
0.02	2013-05-10	<p>Updated text for INSPIRE in section 2.1.4. Added Project Open Data Common Core Metadata Schema, a DCAT Application Profile in section 2.2.4</p>	Team review
0.03	2013-05-10	Corrected listing of WG members in section	Team review
0.04	2013-05-10	Review of the WG members	Team review
0.05	2013-05-10	Addition of the UML Diagram	Team review
0.06	2013-05-13	Update of change type record	Team review

Table of Contents

1. INTRODUCTION	1
1.1. CONTEXT	1
1.2. SCOPE	1
1.3. PROCESS AND METHODOLOGY	2
1.4. STRUCTURE OF THIS DOCUMENT	2
2. RELATED WORK	3
2.1. MODELS FOR DESCRIBING DATASETS OR SIMILAR INFORMATION	3
2.1.1. <i>Asset Description Metadata Schema (ADMS)</i>	3
2.1.2. <i>CERIF for Datasets (CAD)</i>	3
2.1.3. <i>CKAN Dataset Schema</i>	3
2.1.4. <i>INSPIRE Metadata Specification</i>	3
2.1.5. <i>Schema.org</i>	4
2.1.6. <i>Statistical Data and Metadata eXchange (SDMX)</i>	4
2.1.7. <i>Vocabulary of Interlinked Datasets (VOID)</i>	4
2.2. APPLICATION PROFILES.....	5
2.2.1. <i>Austrian Open Government Data Cooperation</i>	5
2.2.2. <i>Norma Técnica de Interoperabilidad de Reutilización de recursos de la información</i>	5
2.2.3. <i>OGD Metadata Structure of govdata.de</i>	5
2.2.4. <i>Project Open Data Common Core Metadata Schema</i>	5
2.2.5. <i>WMO Core Metadata Profile</i>	5
3. USE CASES.....	6
3.1. USER SCENARIO 1 – DATASETS ON TOURIST ACCOMMODATIONS	7
3.2. USER SCENARIO 2 – DATASETS ON EMPLOYMENT RATES, IMMIGRATION, AND IMMIGRATION CONTROL LEGISLATION	7
3.3. USER SCENARIO 3 – CROSS-LINGUAL SEARCH.....	8
3.4. USER SCENARIO 4 – SUBSCRIPTION AND RECOMMENDATIONS	8
3.5. USER SCENARIO 5 – CROSS-BORDER SPATIAL QUERIES FOR DATASETS	9
3.6. USER SCENARIO 6 – GIVE USERS A ROUTE TO CORRECT ERRORS	9
3.7. USER SCENARIO 7 – CREATIONS, UPDATES, AND DELETES OF CATALOG ENTRIES	9
3.8. USER SCENARIO 8 – FEDERATED LEGISLATIVE CATALOGUE AND SEARCH ENGINE	11
4. THE DCAT SPECIFICATION	11
4.1. CURRENT STATUS OF DCAT AT W3C	11
4.2. OVERVIEW OF CLASSES AND PROPERTIES DEFINED IN DCAT	11
4.2.1. <i>Classes</i>	12
4.2.2. <i>Properties</i>	12
5. TERMINOLOGY USED IN THE APPLICATION PROFILE.....	15
6. APPLICATION PROFILE CLASSES	15
6.1. MANDATORY CLASSES.....	16
6.2. OPTIONAL CLASS.....	17
7. APPLICATION PROFILE PROPERTIES PER CLASS	17
7.1. CATALOG	17
7.1.1. <i>Mandatory properties for Catalog</i>	17
7.1.2. <i>Recommended properties for Catalog</i>	17
7.1.3. <i>Optional properties for Catalog</i>	18
7.2. CATALOGRECORD.....	18
7.2.1. <i>Mandatory properties for CatalogRecord</i>	18
7.2.2. <i>Recommended properties for CatalogRecord</i>	18
7.2.3. <i>Optional properties for CatalogRecord</i>	18

7.3.	DATASET	18
7.3.1.	<i>Mandatory properties for Dataset</i>	18
7.3.2.	<i>Recommended properties for Dataset</i>	19
7.3.3.	<i>Optional properties for Dataset</i>	19
7.4.	DISTRIBUTION.....	20
7.4.1.	<i>Mandatory properties for Distribution</i>	20
7.4.2.	<i>Optional properties for Distribution</i>	20
7.5.	AGENT	20
7.5.1.	<i>Mandatory properties for Agent</i>	20
7.6.	CATEGORY SCHEME	21
7.6.1.	<i>Mandatory properties for Category Scheme</i>	21
7.7.	CATEGORY	21
7.7.1.	<i>Mandatory properties for Category</i>	21
7.8.	LICENSE DOCUMENT	21
7.8.1.	<i>Mandatory properties for License Document</i>	21
8.	CONTROLLED VOCABULARIES	21
8.1.	REQUIREMENTS FOR CONTROLLED VOCABULARIES.....	21
8.2.	CONTROLLED VOCABULARIES TO BE USED	22
8.3.	OTHER CONTROLLED VOCABULARIES.....	23
9.	CONFORMANCE STATEMENT.....	23
10.	ACCESSIBILITY AND MULTILINGUAL ASPECTS	23
11.	DEPLOYMENT ISSUES.....	24
11.1.	PUBLISHING LINKED DATA.....	24
11.2.	EXCHANGE OF DATA	24
11.3.	PROVENANCE INFORMATION	25
12.	ACKNOWLEDGEMENTS	25

1. INTRODUCTION

1.1. Context

This document is prepared in the context of Action 1.1 – Improving semantic interoperability in European eGovernment systems¹ of the European Commission's Interoperability for European Public Administrations (ISA) programme². Studies conducted on behalf of the European Commission³ show that businesses and citizens still face difficulties in finding and re-using public sector information. In its communication on Open Data⁴ of 12 December 2011, the European Commission states that the availability of the information in a machine-readable format as well as a thin layer of commonly agreed metadata could facilitate data cross-reference and interoperability and therefore considerably enhance its value for reuse. Much of the public sector information that would benefit from interoperability is published as datasets in data portals. Therefore, an agreement on a common format for data exchange would support the sharing, discovery and re-use of these data.

1.2. Scope

This objective of this work is to define an Application Profile that can be used for the exchange of descriptions of datasets among data portals.

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

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

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

The Application Profile specified in this document is based on the Data Catalog Vocabulary (DCAT) specification⁵ developed under the responsibility of the Government Linked Data Working Group⁶ at W3C. DCAT is an RDF⁷ vocabulary designed to facilitate interoperability between data catalogues published on the Web. Additional classes and properties from other well-known vocabularies are re-used where necessary.

The charter of the Working Group that is developing this Application Profile includes the following objectives:

¹ European Commission. Interoperability for European Public Administrations (ISA). Improving semantic interoperability in European eGovernment systems. http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-laction_en.htm

² European Commission. Interoperability for European Public Administrations (ISA). http://ec.europa.eu/isa/index_en.htm

³ Review of recent studies on PSI reuse and related market developments, Graham Vickery. http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/final_version_study_psi.docx

⁴ European Commission. Communication on Open Data. http://ec.europa.eu/information_society/policy/psi/docs/pdfs/opendata2012/open_data_communication/en.pdf

⁵ W3C. Data Catalog Vocabulary (DCAT). W3C Working Draft, 12 March 2013. <http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/>.

⁶ W3C. Government Linked Data (GLD) Working Group. http://www.w3.org/2011/gld/wiki/Main_Page

⁷ W3C. Resource Description Framework (RDF). <http://www.w3.org/RDF/>

- (1) Identify the essential elements and attributes of DCAT in the European context;
- (2) Identify the controlled vocabularies to be used in the European context; and
- (3) Identify the strict minimum description metadata to be exchanged between data portals in Europe.

The work does not cover implementation issues like mechanisms for exchange of data and expected behaviour of systems implementing the Application Profile other than what is defined in the Conformance Statement in chapter 9.

The Application Profile is intended to facilitate data exchange and therefore the classes and properties defined in this document are only relevant for the data to be exchanged; there are no requirements for communicating systems to implement specific technical environments. The only requirement is that the systems can export and import data in RDF in conformance with this Application Profile.

1.3. Process and methodology

This work is conducted according to a process and methodology⁸ 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.

1.4. Structure of this document

This document consists of the following chapters.

- In chapter 2, a number of related activities are identified.
- Chapter 3 defines the main use case that drives the specification of the Application Profile, namely the exchange of information about data catalogues and datasets among data portals in Europe.
- Chapter 4 contains the base specification of the Data Catalogue Vocabulary on which the Application Profile will be based.
- In chapter 5, the terminology used in the specification of the Application Profile sections is introduced.
- The classes defined for the Application Profile are identified in chapter 6.
- Chapter 7 lists the mandatory and optional properties of those classes used in the Application Profile.
- In chapter 8, controlled vocabularies are proposed for use as value sets for a number of properties.
- Chapter 9 contains the Conformance statement for this Application Profile.
- Accessibility and multilingual issues are addressed in chapter 10.
- In chapter 11, a number of pointers are given that may be helpful for implementation of the Application Profile in a Linked Data environment.
- Finally, acknowledgements related to the development of this Application Profile are contained in chapter 12.

⁸ European Commission. Joinup. Process and Methodology for Developing Core Vocabularies.
<http://joinup.ec.europa.eu/library/document/isa-deliverable-process-and-methodology-developing-core-vocabularies>

2. RELATED WORK

2.1. Models for describing datasets or similar information

2.1.1. Asset Description Metadata Schema (ADMS)

The Asset Description Metadata Schema (ADMS)⁹ is a vocabulary to describe interoperability assets making it possible for ICT developers to discover and reuse interoperability assets. The ADMS namespace document¹⁰ is published by W3C. The class of ADMS Asset is modelled as a subclass of DCAT Dataset.

2.1.2. CERIF for Datasets (C4D)

CERIF¹¹ is an EU recommendation to Member states that defines a data model and XML interchange format for interoperability of research information, maintained by EuroCRIS¹² and used by more than 150 institutional systems across Europe and 10 national systems.

The overall aim of CERIF for Datasets (C4D)¹³ is to develop a framework for incorporating metadata into CERIF such that research organisations and researchers can better discover and make use of existing and future research datasets, wherever they may be held.

2.1.3. CKAN Dataset Schema

The Comprehensive Knowledge Archive Network (CKAN)¹⁴ is a Web-based open source data management system for the distribution of data maintained by the Open Knowledge Foundation¹⁵. The Dataset¹⁶ is the central domain object in the CKAN Domain Model¹⁷.

2.1.4. INSPIRE Metadata Specification

INSPIRE¹⁸ is a Directive of the European Parliament and of the Council aiming to establish a EU-wide spatial data infrastructure to give cross-border access to information that can be used to support EU environmental policies, as well as other policies or activities having an impact on the environment. The actual scope of this information corresponds to 34 environmental themes, covering also areas having cross-sector relevance - e.g., addresses, buildings, population distribution and demography.

In order to ensure cross-border interoperability of data infrastructures operated by EU Member States, INSPIRE sets out a framework based on common specifications

⁹ European Commission. Joinup. Asset Description Metadata Schema (ADMS), version 1.00. 18 April 2012.

<https://joinup.ec.europa.eu/asset/adms/release/100>

¹⁰ W3C. Asset Description Metadata Schema (ADMS). Namespace Document 25 June 2012. <http://www.w3.org/ns/adms>

¹¹ euroCRIS. CERIF Introduction. <http://www.eurocris.org/Index.php?page=CERIFintroduction&t=1>

¹² EuroCRIS Current Research Information Systems, the European Organisation for International Research Information. <http://www.eurocris.org/>

¹³ JISC. CERIF for Datasets (C4D). Delivery scheduled 31 March 2013.

http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/infrastructure/c4d.aspx

¹⁴ CKAN. <http://ckan.org/>

¹⁵ Open Knowledge Foundation. <http://okfn.org/>

¹⁶ CKAN Dataset Schema. <http://docs.ckan.org/en/ckan-1.8/domain-model-dataset.html>

¹⁷ CKAN Domain Model. <http://docs.ckan.org/en/ckan-1.8/domain-model.html>

¹⁸ European Commission. INSPIRE Directive. <http://inspire.ec.europa.eu/>

for metadata, data, network services, data and service sharing, monitoring and reporting. Such specifications consist of a set of implementing rules, along with the corresponding technical guidelines, defined by a regulatory committee composed of representatives of both EU Member States and EU bodies and institutions. The INSPIRE Metadata Implementing Rules¹⁹ include rules for the description of datasets.

2.1.5. *Schema.org*

Schema.org²⁰ is an activity that provides a collection of schemas, i.e., html tags, that webmasters can use to markup their pages in ways recognized by major search providers. Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results, making it easier for people to find the right web pages.

The type hierarchy²¹ includes DataCatalog (a collection of datasets), Dataset (a body of structured information describing some topic(s) of interest) and DataDownload (a dataset in downloadable form) which correspond roughly to Catalog, Dataset and Distribution in DCAT.

2.1.6. *Statistical Data and Metadata eXchange (SDMX)*

Statistical Data and Metadata eXchange (SDMX)²² is an initiative to foster standards for the exchange of statistical information. The specifications include an information model, XML formats and schemas and an UN/EDIFACT format. In addition to defining general descriptors for datasets, SDMX focuses on description of the data and the data structures within datasets. SDMX was published as an International Standard ISO 17369:2013²³.

2.1.7. *Vocabulary of Interlinked Datasets (VoID)*

The Vocabulary of Interlinked Datasets (VoID)²⁴ is an RDF Schema vocabulary for expressing metadata about RDF datasets. It is intended as a bridge between the publishers and users of RDF data, with applications ranging from data discovery to cataloguing and archiving of datasets.

VoID specifies descriptors for the dataset (using the Dublin Core Metadata Terms²⁵), the methods by which the data can be accessed, the schema and internal structure of the data in the dataset, and the links between datasets.

¹⁹ European Commission. Joint Research Centre. INSPIRE Metadata Implementing Rules. 2010-06-16. http://inspire.jrc.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf

²⁰ Schema.org. <http://schema.org/>

²¹ Schema.org. The Type Hierarchy. <http://schema.org/docs/full.html>

²² Statistical Data and Metadata eXchange (SDMX). <http://sdmx.org/>

²³ ISO 17369:2013. Statistical data and metadata exchange (SDMX). http://www.iso.org/iso/catalogue_detail.htm?csnumber=52500

²⁴ W3C. Describing Linked Datasets with the VoID Vocabulary. <http://www.w3.org/TR/void/>

²⁵ Dublin Core Metadata Initiative. DCMI Metadata Terms. <http://dublincore.org/documents/dcmi-terms/>

2.2. Application Profiles

2.2.1. Austrian Open Government Data Cooperation

The Austrian Open Government Data Cooperation²⁶ has produced the DCAT Application Profile called "OGD Metadaten – 2.1"²⁷.

2.2.2. Norma Técnica de Interoperabilidad de Reutilización de recursos de la información

The "Norma Técnica de Interoperabilidad de Reutilización de recursos de la información"²⁸ is an interoperability specification (in Spanish) based on the DCAT vocabulary that harmonises how Spanish public administrations describe datasets or more generally public sector information (PSI).

2.2.3. OGD Metadata Structure of govdata.de

The OGD Metadata Structure²⁹ of govdata.de is a CKAN Profile. It is written as a JSON schema document. It has controlled vocabularies for topics and licenses.

2.2.4. Project Open Data Common Core Metadata Schema

Project Open Data³⁰ is an initiative of the US White House that provides a collection of code, tools, and case studies to help agencies adopt the US Open Data Policy and unlock the potential of government data. Project Open Data will evolve over time as a community resource to facilitate broader adoption of open data practices in government.

A Common Core Metadata Schema³¹ is provided defining 'required', 'required-if-applicable' and 'expanded' fields for the description of Datasets, based on DCAT.

2.2.5. WMO Core Metadata Profile

The WMO Core Metadata Profile³² of the World Meteorological Organization³³ is a profile for use in applications related to weather and climate of the ISO 191xx family of standards developed under responsibility of ISO/TC 211 Geographic information/Geomatics³⁴.

²⁶ Austrian Open Government Data Cooperation <http://reference.e-government.gv.at/Open-Government-Data.2771.0.html>

²⁷ OGD Metadaten – 2.1 <http://reference.e-government.gv.at/Veroeffentlichte-Informationen.2774.0.html>

²⁸ Spain. Ministerio de Hacienda y Administraciones Públicas. Boletín Oficial del Estado. Norma Técnica de Interoperabilidad de Reutilización de recursos de la información. 4 March 2013. <http://www.minhap.gob.es/Documentacion/Publico/NormativaDoctrina/Administracion%20electronica/Resoluci%C3%B3n%2019%20de%20febrero%20de%202013.pdf>

²⁹ OGD Metadata Structure of govdata.de http://htmlpreview.github.io/?https://github.com/fraunhoferfokus/ogd-metadata/blob/master/OGPD_JSON_Schema.html

³⁰ Project Open Data. <http://project-open-data.github.io/>

³¹ Project Open Data. Common Core Metadata Schema. <http://project-open-data.github.io/schema/>

³² World Meteorological Organization. WMO Core Metadata Profile version 1.2. 12 November 2010. http://www.wmo.int/pages/prog/www/WIS/wiswiki/tiki-download_wiki_attachment.php?attId=456&page=ipetmdiPackage&download=y

³³ World Meteorological Organization. <http://www.wmo.int/>

³⁴ ISO/TC 211 Geographic information/Geomatics. <http://www.isotc211.org/Outreach/Overview/Overview.htm>

3. USE CASES

The basic use case that this specification intends to enable is a cross-data portal search for data sets. This can be achieved by the exchange of descriptions of data sets among data portals. The basic use case involves the following actors and systems:

- **Data providers:** Data providers include a description of their datasets on one or more data portals, so that the data sets can be more easily found.
- **Data portals:** Data portals maintain a data catalogue including a collection of datasets made available by data publishers. Data portals make the description metadata of the datasets in their collection freely available to third parties. In addition, data portals may also make collections of relevant datasets of other data portals searchable via their user interface. For enhanced interoperability, the description metadata adheres to the specifications of the DCAT Application Profile.
- **Metadata Brokers:** Metadata Brokers facilitate the exchange of description metadata between data portals by ensuring conformance to the DCAT Application Profile. They provide metadata harvesting, transformation, validation, harmonization, and publication services. The Open Data Support³⁵ project funded by the European Commission will operate a Metadata Broker service for data portals in Europe.
- **Data Consumers:** Users (data consumers) use the data portal of their choice to search through various collections of data sets from a single point of access. The data portal allows the user to explore [FRSAD – Functional Requirements for Subject Authority Records³⁶], find, identify and select [FRBR – Functional Requirements for Bibliographic Records³⁷] data sets coming from different EU Member States, different portals and different organisations. Data consumers could also be systems (machines).

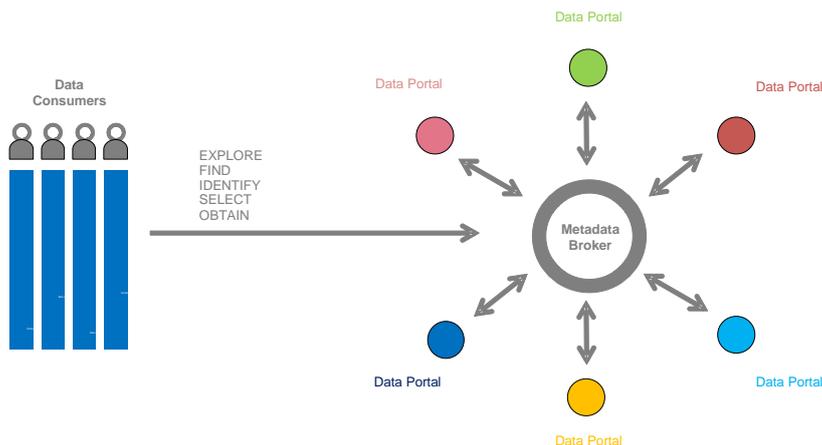


Figure 1 - Basic use case: enable a search for datasets across various data portals

³⁵ Open Data Support. <https://joinup.ec.europa.eu/node/62928>

³⁶ IFLA. Functional Requirements for Subject Authority Data (FRSAD). <http://www.ifla.org/en/node/1297>

³⁷ IFLA. Functional Requirements for Bibliographic Records. <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

3.1. User Scenario 1 – Datasets on tourist accommodations

Tomasz works for a Polish tourist agency and is looking for data sets on tourist accommodations in European cities. His employer wants to develop a travel app to expand its service offerings. Unfortunately, the data sets Tomasz is looking for are made available by various public administrations in Europe which Tomasz is not aware of.

- **Without a DCAT Application Profile:** There are several data portals in Europe that have in their collection data sets about tourist accommodations. Unfortunately, data portals do not exchange their collections of data sets with each other. Therefore, Tomasz cannot do anything else but to search on as many data portals as he can find. In his search he is faced with a variety of user interfaces, description metadata, languages, and classification schemas. It takes Tomasz considerable effort to find suitable datasets.
- **With a DCAT Application Profile:** Data portals exchange description metadata of their own collections using a common metadata vocabulary, and common controlled vocabularies. They are supported by the services of a Metadata Broker. Consequently, Tomasz can search for data sets on a data portal that he is familiar with. He can explore the datasets available in different collections using the preferred user interface, in his own language, using consistent multilingual classification schemes, machine-translated or human-translated metadata values and filtering methods. The description metadata in the search result provide him with sufficient information to identify and select suitable datasets. To obtain the data sets, Tomasz is first directed to the original data portal on which the dataset is listed. From there, he finds further information on how to obtain the dataset.

3.2. User Scenario 2 – Datasets on employment rates, immigration, and immigration control legislation

Julie works for a university and is looking for data sets on migration in the European Union. She wants to carry out a study to analyse the evolution of migration flows from 1950 to 2013, as compared to the variation of employment rate in different Member States. Therefore, she will not only need to look for statistics at national and European level but also at legislation on immigration control. Migration data sets and related legislation are made available by both national and European Union public administrations at different levels in a distributed environment, which Julie is not aware of. The same holds for datasets on employment rates, which are usually made available by institutes of statistics as well as by public administrations and labour unions.

- **Without a DCAT Application profile:** there are several statistics available at national and European level on migration. Their interpretation needs to take into account the evolution of relevant legislation, in order to correctly analyse these trends. Furthermore, the way statistics are gathered may vary depending on the chosen criteria. Without DCAT this search is very cumbersome and takes time: Julie has to identify relevant legislative datasets reporting legislation on the specific matter of immigration control in

the European Union and in different Member States, as well as datasets on employment rates, distributed among several actors. Once Julie has identified such datasets, nevertheless the management of such data is difficult for the variety of user interfaces, metadata and languages.

- **With a DCAT application Profile:** with DCAT as a common metadata vocabulary describing datasets, and by the support of a Metadata Broker service, Julie is able to query such service as unique point of access to identify relevant datasets on immigration control legislation, as well as about statistics on migration and the variation of employment rates in EU and different Member States in specific periods of time. Starting from such information Julie can easily access to the different datasets, select the information of interest, and collect such information. This can be the starting point to develop facilities for data transformation into a common language allowing Julie to mash data up and visualize the variations of employment rates in different geographical EU regions, as well as to compare such data with the legislation on immigration control in force in each specific country and at EU level.

3.3. User Scenario 3 – Cross-lingual search

Belgium has 3 official languages (NL, FR and DE), and often data owners want to add English as well. Of course, metadata (categories, regions, ...) must be searchable in several languages. Our requirements would be to have:

- **Multilingual descriptions:** a short description is available in 4 languages
- **Multilingual links:** link(s) to webpages with more information are available in 3 languages
- **Multilingual datasets:** the dataset itself is only available in 2 languages

It is often helpful that a developer searching for a dataset gets all this info in a single search request and/or that someone who is looking for a dataset in language X can find out that the dataset may (only) exist in language Y.

3.4. User Scenario 4 – Subscription and recommendations

Katarzyna works as a journalist for a newspaper in the Czech Republic and wants to be informed of updates and new datasets related to government spending in her country and neighbouring countries. She has set up the following:

- **Subscriptions:** On her preferred data portal, Katarzyna subscribes to existing public spending datasets of the Czech, Hungarian, and Slovak national governments. She receives an e-mail notification whenever a new distribution is known to be available for these datasets.
- **Recommendation:** On her preferred data portal, Katarzyna indicates that she is interested in datasets related to government spending in her region and neighbouring countries. Whenever a new dataset is known to be available, the system sends her an e-mail notification based on her pre-set preferences.

3.5. User Scenario 5 – Cross-border spatial queries for datasets

Pavel works for an environmental agency in Croatia. He wants to obtain an overview of upstream industrial activity along the river Danube in Germany, Austria, Slovakia, and Hungary.

Pavel navigates to the data portal of his preference (e.g. the INSPIRE geoportal³⁸ or a Czech data portal), enters a search keyword 'industry' in Czech, draws the bounding box of the area of interest to him and is able to retrieve multiple datasets that originate from different countries and that have been catalogued on different data portals with descriptions in different languages. Pavel can further refine his search by adjusting geographic coverage or filtering on the theme, spatial coverage (geographic names), temporal coverage, license, etc. of the datasets in the search results. The search results may contain a graphical representation of the spatial coverage on a map.

3.6. User Scenario 6 – Give users a route to correct errors

Whenever someone accesses a dataset, it may happen that the user notes that there is an error in the data set. In such cases, the user may want to contact the maintainer of the dataset to suggest correction of the error. To enable the user to do this, contact information of the maintainer is made available which the user can use to send an e-mail with an error notification.

3.7. User Scenario 7 – Creations, updates, and deletes of catalog entries

When data portals exchange description metadata, they need a mechanism to keep the exchanged metadata up-to-date. Otherwise, outdated description metadata might pollute the "federation of data portals". For example, without a proper mechanism, deleted datasets continue to be listed on the websites of aggregators.

This mechanism can be based on the exchange of catalog records (A) or on the exchange of an entire snapshot (B).

Mechanism A. Exchange based on catalog records: A set of catalog records that have been created, updated, or deleted after a specific time interval – typically the last update period – is exchanged between a Data Portal and a Metadata Broker. This happens in the following steps:

1. **Recordkeeping by Data Portal:** The Data Portal keeps track of catalog records that represent the latest create, update, and delete transactions to its metadata;
2. **Exchange (push or pull):** Periodically, the Data Portal pushes the catalog records that have been created, updated, or deleted to the Metadata Broker. Alternatively, the Metadata Broker periodically pulls (metadata harvesting) the metadata records that have been created, updated, or deleted from the Data Portal.

³⁸ European Commission. INSPIRE Geoportal. <http://inspire-geoportal.ec.europa.eu/>

3. **Update by Metadata Broker:** The Metadata Broker updates its own metadata to reflect the changes indicated in the catalog records.
 - **Created records:** It will create the metadata for all catalog records that have been created. For example, if a new dataset was added to the collection of the Data Portal, the Metadata Broker will incorporate its description metadata;
 - **Updated records:** It will reflect updates to the metadata for all catalog records that indicate an update of metadata. For example, if the description metadata of a dataset was updated on the Data Portal, the Metadata Broker will reflect all changes;
 - **Delete records:** It will delete the metadata for all catalog records that indicate a deletion of metadata. For example, if a data set was removed from the collection of a Data Portal, the Metadata Broker will reflect this.
4. **Recordkeeping by Metadata Broker:** The Metadata Broker uses the same catalog records as the Data Portal. In turn, he can offer a CatalogRecord-based exchange of metadata.

Mechanism B. Snapshot-based exchange: A metadata snapshot is exchanged between a Data Portal and a Metadata Broker that contains all metadata exactly as it appears at a specific point in time.

1. **No recordkeeping by the Data Portal:** The Data Portal does not (need to) keep track of catalog records.
2. **Exchange (push or pull):** Periodically, the Data Portal pushes a snapshot of all its metadata to the Metadata Broker. Alternatively, the Metadata Broker pulls (metadata harvesting) a snapshot from the Data Portal.
3. **Update by Metadata Broker:** The Metadata Broker updates its own metadata but also incorporates catalog records to reflect creates, updates, and deletes to the metadata. The latter can be achieved if the Metadata Brokers compares the snapshot with a previous snapshot for the Data Portal.
 - **Unchanged metadata:** The Metadata Broker does not update the metadata nor the corresponding catalog records. For example, if a description of a dataset remains unchanged between the current and the previous snapshot, no updates are needed.
 - **Created metadata:** The Metadata Broker adds metadata which has been added to the snapshot and also creates a catalog record to reflect this. For example, if a description of a dataset was added to the current snapshot that was not present in the previous snapshot, the Metadata Broker will also incorporate this description metadata and it will create a catalog record to reflect the creation.
 - **Updated metadata:** The Metadata Broker updates metadata which has been updated and also updates the modification date of the catalog record to reflect this. For example, if the title of a dataset is updated, the Metadata Broker will apply this update and update the modification date of the corresponding catalog record to reflect this.
 - **Deleted metadata:** By comparing the snapshot with a previous snapshot, the Metadata Broker detects that some metadata has been removed, it also remove the metadata, but leave a catalog record to reflect this deletion. For example, if a dataset is removed from the

collection of a Data Portal, the Metadata Broker will delete the Dataset and include information about the "deleted entry" in its catalog records.

4. **Recordkeeping by Metadata Broker:** The Metadata Broker now can offer a CatalogRecord-based exchange of metadata.

3.8. User Scenario 8 – Federated Legislative Catalogue and Search Engine

Each Region of Italy (22, included autonomous Provinces) owns a database of the local legislation, published through a web portal. Users willing to perform a search on all these legislative databases, need to "travel" across the 22 portals, in order to find the acts of interest. Search interfaces and criteria vary among portals along with the way in which search results are presented. A list of these databases is provided here: <http://www.normattiva.it/static/mappa.html>. In order to ease the burden of searching acts across all these portals, a federated search engine is under development. The ideas underlying the design of this engine are: each Region publishes a catalogue of all its legislation at a known URL; the catalogue contains the list of the legislative references to all the acts of a given Region, along with URLs to each act; a federated indexer - exploiting the catalogues - crawls the text-only version of all acts in order to build a cross-regional federated index, on top of which a federated search engine permits to perform searches. In our view, the legislative catalogue of each Region could be described and published using a DCAT application profile. This profile could also be applied to national and European legislations.

4. THE DCAT SPECIFICATION

4.1. Current status of DCAT at W3C

The specification of the Data Catalog Vocabulary (DCAT), version 12 March 2013, was published by the Government Linked Data Working Group as a Last Call Working Draft with a deadline for comments of 08 April 2013.

During the Last Call comment period comments were made that are being discussed by the GLD Working Group. If changes are applied to the DCAT specification, the description of the classes and properties will be updated in further versions of the proposed Application Profile.

4.2. Overview of classes and properties defined in DCAT

Please note that the information in the following sections has been copied from the current version of the DCAT specification. The information is provided here for easier reference in the discussions of the DCAT AP Working Group. In the final specification of the Application Profile, the detailed information about the base specification of DCAT will be removed and replaced by a reference to the specification on the W3C Website.

4.2.1. Classes

Namespaces prefixes used in Table 1 below:

- dcat: <http://www.w3.org/ns/dcat>
- foaf: <http://xmlns.com/foaf/0.1/>
- skos: <http://www.w3.org/2004/02/skos/core>

Table 1: DCAT Classes

Class name	Usage note	URI	Reference
Catalog	Typically, a web-based data catalog is represented as a single instance of this class.	dcat:Catalog	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog
Catalog record	This class is optional and not all catalogs will use it. It exists for catalogs where a distinction is made between metadata about a dataset and metadata about the dataset's entry in the catalog. For example, the publication date property of the dataset reflects the date when the information was originally made available by the publishing agency, while the publication date of the catalog record is the date when the dataset was added to the catalog. In cases where both dates differ, or where only the latter is known, the publication date should only be specified for the catalog record.	dcat:CatalogRecord	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog-record
Dataset	This class represents the actual dataset as published by the dataset publisher. In cases where a distinction between the actual dataset and its entry in the catalog is necessary (because metadata such as modification date and maintainer might differ), the catalog record class can be used for the latter.	dcat:Dataset	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-dataset
Distribution	This represents a general availability of a dataset it implies no information about the actual access method of the data, i.e. whether it is a direct download, API, or some through Web page. The use of dcat:downloadURL property indicates directly downloadable distributions.	dcat:Distribution	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-distribution
Category and category scheme	It is necessary to use either skos:inScheme or skos:topConceptOf on every skos:Concept otherwise it's not clear which concept scheme they belong to.	skos:ConceptScheme , skos:Concept	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-category-and-category-scheme
Organization/ Person	FOAF provides sufficient properties to describe these entities.	foaf:Person , foaf:Organization	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-organization-person

4.2.2. Properties

Namespaces prefixes used in Table 2 below:

- dcat: <http://www.w3.org/ns/dcat#>
- dct: <http://purl.org/dc/terms/>
- dctype: <http://purl.org/dc/dcmitype/>
- foaf: <http://xmlns.com/foaf/0.1/>
- rdfs: <http://www.w3.org/2000/01/rdf-schema#>
- skos: <http://www.w3.org/2004/02/skos/core#>
- xsd: <http://www.w3.org/2001/XMLSchema#>

Table 2: DCAT properties, alphabetically per class

Class	Property	URI	Range	Usage note
dcat:Catalog	catalog record	dcat:record	dcat:CatalogRecord	A catalog record that is part of the catalog.
dcat:Catalog	dataset	dcat:dataset	dcat:Dataset	A dataset that is part of the catalog.
dcat:Catalog	description	dct:description	rdfs:Literal	Free-text account of the catalog.
dcat:Catalog	homepage	foaf:homepage	foaf:Document	The homepage of the catalog.
dcat:Catalog	language	dct:language	rdfs:Resource	The language of the catalog. This refers to the language used in the textual metadata describing titles, descriptions, etc. of the datasets in the catalog.
dcat:Catalog	license	dct:license	dctype:LicenseDocument	This describes the license under which the catalog can be used/reused and not the datasets. Even if the license of the catalog applies to all of its datasets and distributions, it should be replicated on each distribution.
dcat:Catalog	publisher	dct:publisher	foaf:Agent	The entity responsible for making the catalog online.
dcat:Catalog	release date	dct:issued	rdfs:Literal typed as xsd:date	Date of formal issuance (e.g., publication) of the catalog.
dcat:Catalog	spatial/ geographic	dct:spatial	dct:Location	The geographical area covered by the catalog.
dcat:Catalog	themes	dcat:themeTaxonomy	skos:ConceptScheme	The knowledge organization system (KOS) used to classify catalog's datasets.
dcat:Catalog	title	dct:title	rdfs:Literal	A name given to the catalog.
dcat:Catalog	update/ modification date	dct:modified	rdfs:Literal typed as xsd:date	Most recent date on which the catalog was changed, updated or modified.
dcat:CatalogRecord	description	dct:description	rdfs:Literal	Free-text account of the record.
dcat:CatalogRecord	listing date	dct:issued	rdfs:Literal typed as xsd:date	The date of listing the corresponding dataset in the catalog.
dcat:CatalogRecord	primary topic	foaf:primaryTopic	dcat:Dataset	Links the catalog record to the dcat:Dataset resource described in the record.
dcat:CatalogRecord	title	dct:title	rdfs:Literal	A name given to the record.
dcat:CatalogRecord	update/ modification date	dct:modified	rdfs:Literal typed as xsd:date	Most recent date on which the catalog entry was changed, updated or modified.
dcat:Dataset	dataset distribution	dcat:distribution	dcat:Distribution	Connects a dataset to its available distributions.
dcat:Dataset	description	dct:description	rdfs:Literal	Free-text account of the dataset.
dcat:Dataset	frequency	dct:accrualPeriodicity	dct:Frequency	The frequency at which dataset is published.

DCAT Application Profile for Data Portals in Europe

Class	Property	URI	Range	Usage note
dc:Dataset	identifier	dct:identifier	rdfs:Literal	The identifier might be used as part of the URI of the dataset, but still having it represented explicitly is useful.
dc:Dataset	keyword/tag	dc:keyword	rdfs:Literal	A keyword or tag describing the dataset.
dc:Dataset	landing page	dc:landingPage	foaf:Document	A Web page that can be navigated to in a Web browser to gain access to the dataset, its distributions and/or additional information.
dc:Dataset	language	dct:language	rdfs:Resource	The language of the dataset.
dc:Dataset	publisher	dct:publisher	foaf:Agent	An entity responsible for making the dataset available.
dc:Dataset	release date	dct:issued	rdfs:Literal typed as xsd:date	Date of formal issuance (e.g., publication) of the dataset.
dc:Dataset	spatial/ geographical coverage	dct:spatial	dct:Location	Spatial coverage of the dataset.
dc:Dataset	temporal coverage	dct:temporal	dct:PeriodOfTime	The temporal period that the dataset covers.
dc:Dataset	theme/ category	dc:theme, subproperty of dct:subject	skos:Concept	The main category of the dataset. A dataset can have multiple themes.
dc:Dataset	title	dct:title	rdfs:Literal	A name given to the dataset.
dc:Dataset	update/ modification date	dct:modified	rdfs:Literal typed as xsd:date	Most recent date on which the dataset was changed, updated or modified.
dc:Distribution	access URL	dc:accessURL	rdfs:Resource	Could be any kind of URL that gives access to a distribution of the dataset. E.g. landing page, download, feed URL, SPARQL endpoint. Use when your catalog does not have information on which it is or when it is definitely not a download.
dc:Distribution	byteSize	dc:byteSize	rdfs:Literal typed as xsd:decimal	The size of a distribution in bytes.
dc:Distribution	description	dct:description	rdfs:Literal	Free-text account of the distribution.
dc:Distribution	download URL	dc:downloadURL	rdfs:Resource	This is a direct link to a downloadable file in a given format. E.g. CSV file or RDF file. The format is described by the distribution's dc:format and/or dc:mediaType
dc:Distribution	license	dct:license	dct:LicenseDocument	The license under which the distribution is made available.
dc:Distribution	release date	dct:issued	rdfs:Literal typed as xsd:date	Date of formal issuance (e.g., publication) of the distribution.
dc:Distribution	title	dct:title	rdfs:Literal	A name given to the distribution.

Class	Property	URI	Range	Usage note
dcat:Distribution	update/ modification date	dct:modified	rdfs:Literal typed as xsd:date	Most recent date on which the distribution was changed, updated or modified.
dcat:Distribution	media type	dcat:mediaType, subproperty of dct:format	dct:MediaTypeOrExtent	This property should be used when the media type of the distribution is defined in IANA, otherwise dct:format may be used with different values.
dcat:Distribution	format	dct:format	dct:MediaTypeOrExtent	The file format of the distribution.

5. TERMINOLOGY USED IN THE APPLICATION PROFILE

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.
- **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 RFC2119³⁹.

Note: the obligation for a receiver to process information for mandatory, recommended or optional elements does not include an obligation to make such information searchable or visible to the users of the system. How or even whether the information is used is entirely at the discretion of the application at the receiver's end.

6. APPLICATION PROFILE CLASSES

Figure 2 shows a UML diagram of all classes and properties included in the DCAT Application Profile. In the following sections the classes and properties are defined.

³⁹ IETF. RFC2119. Key words for use in RFCs to Indicate Requirement Levels. <http://www.ietf.org/rfc/rfc2119.txt>

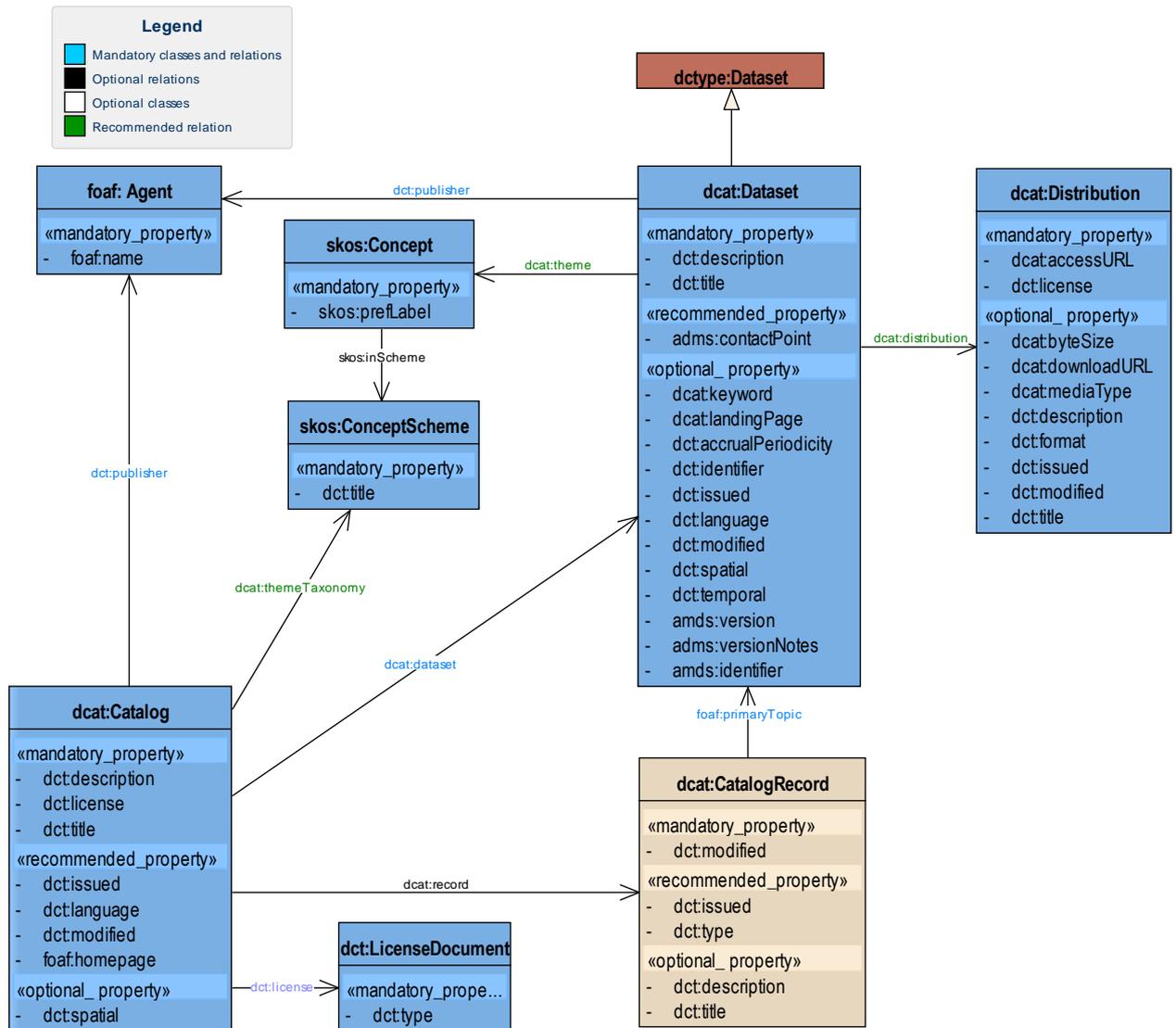


Figure 2 - DCAT Application Profile UML Diagram

6.1. Mandatory Classes

Class name	Usage note for the Application Profile	URI	Reference
Agent	An entity that is associated with catalogues and/or datasets. If the Agent is an organisation, the use of the organisation ontology is recommended.	foaf:Agent	http://xmlns.com/foaf/spec/#term-Agent , http://www.w3.org/TR/vocab-org/
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
Catalog	A catalogue or repository that hosts the datasets being described.	dcat:Catalog	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog

Class name	Usage note for the Application Profile	URI	Reference
Dataset	A conceptual entity that represents the information published.	dcat:Dataset	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-dataset
Distribution	A physical embodiment of the dataset in a particular format.	dcat:Distribution	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-distribution
License 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

6.2. Optional Class

Class name	Reason for exclusion	URI	Reference
Catalog record	This class enables the identification and description of catalogue records.	dcat:CatalogRecord	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog-record

7. APPLICATION PROFILE PROPERTIES PER CLASS

7.1. Catalog

7.1.1. Mandatory properties for Catalog

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.	1..n
license	dct:license	dct:LicenseDocument	This property refers to the license under which the catalogue can be used or reused.	1..1
publisher	dct:publisher	foaf:Agent	This property refers to an entity 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

7.1.2. Recommended properties for Catalog

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	rdfs:Resource	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
release date	dct:issued	rdfs:Literal typed as xsd:date	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 (KOS) used to classify the catalogue's datasets.	0..n
update/ modification date	dct:modified	rdfs:Literal typed as xsd:date	This property contains the most recent date on which the catalogue was changed or modified.	0..1

7.1.3. Optional properties for Catalog

Property	URI	Range	Usage note	Card.
spatial/ geographic	dct:spatial	dct:Location	This property refers to a geographical area covered by the catalogue. It can be represented using a controlled vocabulary or with geographic coordinates. In the latter case, the use of the Core Location Vocabulary ⁴⁰ is recommended.	0..n

7.2. CatalogRecord

7.2.1. Mandatory properties for CatalogRecord

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	This property contains the most recent date on which the catalogue entry was changed or modified.	1..1

7.2.2. Recommended properties for CatalogRecord

Property	URI	Range	Usage note	Card.
listing date	dct:issued	rdfs:Literal typed as xsd:date	This property contains the date on which the description of the Dataset was included in the catalogue.	0..1
change type	adms:status	skos:Concept	The type of the <i>latest</i> revision of a dataset's entry in the catalog. MUST take one of the values :created, :updated or :deleted depending on whether this <i>latest</i> revision is as a result of a creation, update or deletion.	0..1

7.2.3. Optional properties for CatalogRecord

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
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.	0..n

7.3. Dataset

7.3.1. 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

⁴⁰ European Commission. Joinup. Core Location Vocabulary. https://joinup.ec.europa.eu/asset/core_location/description

Property	URI	Range	Usage note	Card
publisher	dct:publisher	foaf:Organization	This property refers to an organisation responsible for making the dataset available.	1..1
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

7.3.2. Recommended properties for Dataset

Property	URI	Range	Usage note	Card
contact point	adms:contactPoint	v:VCard	This property contains contact information that can be used for flagging errors in the dataset or sending comments	0..n
dataset distribution	dcat:distribution	dcat:Distribution	This property links the dataset to an available distribution.	0..n
theme/ category	dcat:theme, subproperty of dct:subject	skos:Concept	This property refers to a category of the dataset. A dataset can have multiple themes.	0..n

7.3.3. Optional properties for Dataset

Property	URI	Range	Usage note	Card.
frequency	dct:accrualPeriodicity	dct:Frequency	This property refers to the frequency at which dataset is published.	0..1
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 Catalog	0..n
keyword/ tag	dcat:keyword	rdfs:Literal	This property contains a keyword or tag describing the dataset.	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.	0..1
language	dct:language	rdfs:Resource	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, e.g. such as MAST/ADS ⁴¹ , DOI ⁴² , EZID ⁴³ or W3ID ⁴⁴ .	
release date	dct:issued	rdfs:Literal typed as xsd:date	This property contains the date of formal issuance (e.g., publication) of the dataset.	0..1
spatial/ geographical coverage	dct:spatial	dct:Location	This property refers to a geographic region that is covered by the dataset. It can be represented using a controlled vocabulary or with geographic coordinates. In the latter case, the use of the Core Location Vocabulary ⁴⁵ is recommended.	0..n
temporal coverage	dct:temporal	dct:PeriodOfTime	This property refers to a temporal period that the dataset covers.	0..n
update/ modification date	dct:modified	rdfs:Literal typed as xsd:date	This property contains the most recent date on which the dataset was changed or modified.	0..1

⁴¹ Mikulski Archive for Space Telescopes (MAST). Referencing Data Sets in Astronomical Literature. http://archive.stsci.edu/pub_dsn.html

⁴² DOI. Digital Object Identifier. <http://www.doi.org/>

⁴³ EZID. <http://n2t.net/ezip>

⁴⁴ W3C Permanent Identifier Community. Permanent Identifiers for the Web. <https://w3id.org/>

⁴⁵ European Commission. Joinup. Core Location Vocabulary. https://joinup.ec.europa.eu/asset/core_location/description

Property	URI	Range	Usage note	Card.
version	adms:version	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.	0..1

7.4. Distribution

7.4.1. 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 that is present at the accessURL may contain information about how to get the dataset if it is not available in digital form.	1..n
license	dct:license	dct:LicenseDocument	This property refers to the license under which the distribution is made available.	1..1

7.4.2. Optional properties for Distribution

Property	URI	Range	Usage note	Card.
byteSize	dcat:byteSize	rdfs:Literal typed as xsd:decimal	This property contains the size of a distribution in bytes.	0..1
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
download URL	dcat:downloadURL	rdfs:Resource	This property contains a URL that is direct link to a downloadable file in a given format.	0..n
format	dct:format	dct:MediaTypeOrExtent	This property refers to the file format of the distribution.	0..1
release date	dct:issued	rdfs:Literal typed as xsd:date	This property contains the date of formal issuance (e.g., publication) 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	This property contains the most recent date on which the distribution was changed or modified.	0..1
media type	dcat:mediaType, subproperty of dct:format	dct:MediaTypeOrExtent	This property refers to the media type of the distribution if this is defined in IANA.	0..1

7.5. Agent

7.5.1. 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

7.6. Category Scheme

7.6.1. 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

7.7. Category

7.7.1. 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

7.8. License Document

7.8.1. Mandatory properties for License Document

Property	URI	Range	Usage note	Card.
license type	dct:type	rdfs:Class	This property refers to a type of license, e.g. indicating 'public domain', or 'royalties required' etc.	1..1

8. CONTROLLED VOCABULARIES

8.1. Requirements for controlled vocabularies

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

Controlled vocabularies SHOULD:

- Be published under an open license
- 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.

8.2. Controlled vocabularies to be used

In the table below, a number of properties are listed with controlled vocabularies that MUST be used for the listed properties.

Property URI	Used for Class	Proposed vocabulary	Vocabulary URI	Usage note
dc:mediaType	Distribution	MDR File types Name Authority List ⁴⁶	http://publications.europa.eu/resource/authority/file-type	
dc:theme	Dataset	EuroVoc domains	http://eurovoc.europa.eu/100142 through 100162	
dc:themeTaxonomy	Catalog	EuroVoc	http://eurovoc.europa.eu/	
dct:accrualPeriodicity	Dataset	Dublin Core Collection Description Frequency Vocabulary ⁴⁷	http://purl.org/cld/freq/	
dct:format	Distribution	MDR File Type Named Authority List	http://publications.europa.eu/resource/authority/file-type	
dct:language	Catalog, Dataset	MDR Languages Named Authority List ⁴⁸	http://publications.europa.eu/resource/authority/language	
dct:publisher	Catalog, Dataset	MDR Corporate bodies Named Authority List ⁴⁹	http://publications.europa.eu/resource/authority/corporate-body	The terms in this vocabulary are only relevant if the organisation is a European institution. In case of other types of organisations, national, regional or local vocabularies should be used.
dct:spatial	Catalog, Dataset	MDR Countries Named Authority List ⁵⁰ , MDR Places Named Authority List ⁵¹	http://publications.europa.eu/resource/authority/country , http://publications.europa.eu/resource/authority/place/	The Countries vocabulary is to be used if the scope of a dataset is limited to a particular country. Additionally, the Places vocabulary is to be used if a dataset is limited to a part of a country.
adms:status	CatalogRecord	ADMS change type vocabulary	http://purl.org/adms/changetype	:created, :updated, :deleted
dct:type	License Document	ADMS license type vocabulary	http://purl.org/adms/licencetype	

⁴⁶ Publications Office of the EU. Metadata Registry. Authorities. File types. <http://publications.europa.eu/mdr/authority/file-type/>

⁴⁷ Dublin Core Metadata Initiative. Dublin Core Collection Description Frequency Vocabulary. <http://dublincore.org/groups/collections/frequency/>

⁴⁸ Publications Office of the EU. Metadata Registry. Authorities. Languages. <http://publications.europa.eu/mdr/authority/language/>

⁴⁹ Publications Office of the EU. Metadata Registry. Authorities. Corporate bodies. <http://publications.europa.eu/mdr/authority/corporate-body/>

⁵⁰ Publications Office of the EU. Metadata Registry. Authorities. Countries. <http://publications.europa.eu/mdr/authority/country/>

⁵¹ Publications Office of the EU. Metadata Registry. Authorities. Places. <http://publications.europa.eu/mdr/authority/place/>

8.3. Other controlled vocabularies

In addition to the proposed common vocabularies in section 8.2, further region or domain-specific vocabularies can be used. 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, the CERIF standard vocabularies⁵², the Dewey Decimal Classification⁵³ and numerous other schemes.

9. CONFORMANCE STATEMENT

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 in section 7.1.1
- Provide information for the mandatory properties specified in section 7.2.1, 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 in section 7.3.1
- Provide descriptions of distributions of datasets in the catalogue, including at least the mandatory properties specified in section 7.4.1
- Provide descriptions of all organisations involved in the descriptions of catalogue and datasets, including at least the mandatory properties specified in section 7.5.1
- 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 in section 7.6.1
- Provide descriptions of all categories involved in the descriptions of datasets in the catalogue, including at least the mandatory properties specified in section 7.7.1

In addition to the mandatory properties, any of the recommended and optional properties defined in chapter 7 **MAY** be provided.

For the properties listed in the table in chapter 8, the associated controlled vocabularies **MUST** be used. Additional controlled vocabularies **MAY** be used.

10. ACCESSIBILITY AND MULTILINGUAL ASPECTS

Accessibility in the context of this Application Profile is limited to information about the technical format of distributions of datasets. The properties `dc: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

⁵² http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.5/CERIF1.5_Semantics.xhtml

⁵³ OCLC. Dewey Summaries as Linked Data. <http://www.oclc.org/dewey/webservices.en.html> and <http://dewey.info/>

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 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 organisation 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, the language tag is mandatory. 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.

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

11. DEPLOYMENT ISSUES

11.1. Publishing Linked Data

As this Application Profile is intended for use in a Linked Data⁵⁴ environment, publishers should consider the recommendations in the W3C Notes "Best Practice Recipes for Publishing RDF Vocabularies"⁵⁵, "Best Practices for Publishing Linked Data"⁵⁶ and the ISA report "10 Rules for Persistent URIs"⁵⁷.

11.2. Exchange of data

While this Application Profile concentrates on the specification of the data format to be used for exchange of information about datasets, in practical situations the communicating partners will need to identify the exchange mechanisms and protocols to be used.

Various approaches may be deployed:

- Harvesting: an aggregator initiates a connection to the data store at a data provider to pull descriptions from the provider's catalogue;

⁵⁴ W3C. Linked Data. <http://www.w3.org/standards/semanticweb/data>

⁵⁵ W3C. Best Practice Recipes for Publishing RDF Vocabularies. <http://www.w3.org/TR/swbp-vocab-pub/>

⁵⁶ W3C. Best Practices for Publishing Linked Data. <https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>

⁵⁷ European Commission. Joinup. 10 Rules for Persistent URIs. <https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>

- File transfer: an aggregator pulls a file with descriptions from the data provider, or the data provider uploads such a file to the aggregator. Such a file is prepared by the data provider as a (partial) export from its catalogue;
- Online maintenance: the data provider maintains the descriptions of its datasets at the aggregator using an online user interface that allows upload, modification and deletion of descriptions.

Various technical specifications can support such mechanisms, such as SPARQL⁵⁸, OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting)⁵⁹, Atom Syndication Format⁶⁰ with ‘tombstones’⁶¹, the Protocol for the Syndication of Resource Descriptions (SDShare)⁶², the Data Catalog Interoperability Protocol⁶³ and others.

11.3. Provenance information

According to the base DCAT specification, the class `CatalogRecord` can be used to capture provenance information about dataset entries in a catalogue. In this Application Profile, the only type of provenance information that is included is the recommended ‘change type’ property `dct:type`, defined in section 7.2.2, which is intended to help determine which descriptions of data sets have been created, updated, or deleted from the data provider’s catalogue.

12. ACKNOWLEDGEMENTS

This work was elaborated by a Working Group under the ISA programme. The Working Group was chaired by **Antonio Carneiro** from the Publications Office. The European Commission was represented by **Vassilios Peristeras** and **Szabolcs Szekacs**. **Makx Dekkers** was the editor of the specification.

The members of the Working Group are enlisted in the following table:

Family name	First name	Organisation	Portal	
Acuña	Antonio	Cabinet Office	data.gov.uk	UK
Alcioto	Nives	CSI Piemonte	dati.piemonte.it	IT
Álvarez Espinar	Martín	W3C Spain Office	Datos.gob.es	ES
Arndt	Adam	Danish Agency for Digitisation	Basic Data, data.digitaliser.dk	DK
Åsnes	Øystein	Difi	data.norge.no	NO
Beyer	Jan-Ole	Federal Ministry of the Interior, Germany	govdata.de	DE
Biro	Peter	Ministry of Finance of the Slovak Republic		SK
Both	Wolfgang	City of Berlin, Dept. for Economics, Technology and Reseach	date.berlin.de	DE

⁵⁸ W3C. SPARQL Query Language for RDF. <http://www.w3.org/TR/rdf-sparql-query/>

⁵⁹ Open Archives Initiative. Protocol for Metadata Harvesting. <http://www.openarchives.org/pmh/>

⁶⁰ IETF. RFC 4287. The Atom Syndication Format. <http://www.ietf.org/rfc/rfc4287>

⁶¹ IETF. RFC6721. The Atom "deleted-entry" Element. <http://tools.ietf.org/html/rfc6721>

⁶² SDShare. A Protocol for the Syndication of Resource Descriptions. <http://www.sdshare.org/spec/sdshare-20120710.html>

⁶³ Data Catalog Interoperability Protocol. <http://spec.datacatalogs.org/>

DCAT Application Profile for Data Portals in Europe

Brasse	Valerie	euroCRIS	eurocris.org	EU
Broomfield	Heather	Difi	data.norge.no	NO
Canabal	José Manuel	Red.es	Datos.gob.es	ES
Caracciolo	Caterina	Food and Agriculture Organization of the United Nations (FAO)	data.fao.org	IT
Castro García-Muñoz	Sonia	Red.es	Datos.gob.es	ES
Combetto	Marco	Informatica Trentina	Trentino Open Data	IT
Cornero	Alessandra	Formez PA	dati.gov.it	IT
Craglia	Massimo	European Commission - Joint Research Centre	Inspire Geoportal	EU
Dragoni	Alex	SciamLab	www.opendatahub.it	IT
Dutkowski	Simon	Fraunhofer FOKUS	govdata.de	DE
Fernández Ruiz	María Jesús	City of Zaragoza	DatosAbiertos.zaragoza.es	ES
Francesconi	Enrico	Institute of Legal Information Theory and Techniques (ITTIG - CNR)	Italian National Research Council	IT
Gutteridge	Christopher	University of Southampton	data.southampton.ac.uk	UK
Hanssens	Bart	Fedict	Data.gov.be	BE
Hermans	Paul	ProXML	Flemish Open Data Portal	BE
Hernández Guerra de Aguilar	Nicolas	Canary Islands Government	UNIFICA, Sielocal	ES
Heus	Pascal	Metadata Technology North-America	http://www.dwbproject.org/	US
Holm	Jeanne	U.S. General Services Administration	Data.gov	US
Houssos	Nikos	euroCRIS	eurocris.org	EU
Ibáñez Pascual	Antonio	Junta de Castilla y León	datosabiertos.jcyl.es	ES
Iglesias	Carlos	Independent Consultant	datos.gob.es, Andalusia Open Data, Open Data Euskadi	ES
Joerg	Brigitte	euroCRIS	eurocris.org	EU
Kaltenböck	Martin	Semantic Web Company (SWC)	LOD2, opendata.ch	AT
Kanellopoulos	Ioannis	European Commission - Joint Research Centre	Inspire Geoportal	EU
Klauser	Dominik	Federal Chancellery of Austria	data.gv.at	AT
Krantz	Peter	The eGovernment Delegation, Peter Kranz AB	OpenGov Sweden Catalog	SE
Koistinen	Kai	National Land Survey of Finland	paikkatietohakemisto	FI
Lacombe	Romain	Etalab	data.gouv.fr	FR
Lalle	Giovanni	Senato della Repubblica	dati.senato.it	IT
Loutas	Nikolaos	PwC Belgium	Open Data Support	BE
Lutz	Michael	European Commission - Joint Research Centre	Inspire Geoportal	EU
Maali	Fadi	DERI		IE
Marchetti	Carlo	Head of Information Systems Development, Senato della Repubblica	dati.senato.it	IT
Marienfeld	Florian	Fraunhofer FOKUS	govdata.de	DE
Maurino	Andrea	University of Milano Bicocca	Municipality of Vigevano - Open Data Portal, dati.lombardia.it	IT
Nativi	Stefano	FP7	EuroGEOSS	IT
Overbeek	Hans	Ministry for the Interior of the Netherlands	data.overheid.nl	NL
Papadopoulos	Thodoris	Greek Prime Minister's Cabinet , Strategic Planning Bureau	data.gov.gr	GR
Perego	Andrea	European Commission -Joint Research Centre	Inspire Geoportal	EU

DCAT Application Profile for Data Portals in Europe

Romain	Pascal	Conseil Général de la Gironde	Ressourcerie Datalocale	FR
Rousseau	Bart	City of Ghent - E-Strategy	Open Data Gent	BE
Ruelle	Charles	Etalab	data.gouv.fr	FR
Sánchez Magento	Alejandra	National Centre of Geographic Information	idee.es	ES
Saponja	Danica	Ministry of Justice and Public Administration of Slovenia		SI
Staromiejski	Enric	Everis	Several European, national and local data portals	ES
Subero	Jose M	opendata.aragon.es	opendata.aragon.es	ES
Thoeve	Thimo	City of Ghent - E-Strategy	Open Data Gent	BE
Van der Waal	Sander	Open Knowledge Foundation	publicdata.eu	UK