DCAT Application Profile for data portals in Europe

Final version

Date: 15/05/2014

Document Metadata

Property	Value
Date	2013-09-02
Status	Final
Version	1.01
Rights	© 2013 European Union
Licence	ISA Open Metadata Licence v1.1, retrievable from https://joinup.ec.europa.eu/category/licence/isa-open-metadata-licence-v11.
Access URL	This specification can be downloaded from the Joinup web site: 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
Draft 1	2013-03-28	First Working Group draft	WG review
Draft 2	2013-04-19	Second Working Group draft	WG review
Draft 3	2013-05-02	Third Working Group draft	Public review
Final draft	2013-07-16	Final Working Group Draft with resolutions of comments received	WG review
Final 0.09	2013-08-06	Final version	Publication
Final 1.00	2013-09-02	Editorial update	Publication
Final 1.01	2013-05-15	Correction of broken links	Publication

Table of Contents

1. INTRODUCTION

1.1. Context

This document is prepared in the context of Action 1.1 – Improving semantic interoperability in European eGovernment systems[1] of the European Commission's Interoperability for European Public Administrations (ISA) programme[2].

Studies conducted on behalf of the European Commission[3] show that businesses and citizens still face difficulties in finding and re-using public sector information. In its communication on Open Data[4] 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, recommended and optional elements to be used for a particular application, as well as recommendations for controlled vocabularies to be used.

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

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

The Application Profile specified in this document is based on the specification of the **Data Catalog Vocabulary** (DCAT)[5] developed under the responsibility of the Government Linked Data Working Group[6] at W3C. The work on DCAT was initiated[7]

[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/1-1action_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] Fadi Maali, Richard Cyganiak, Vassilios Peristeras: Enabling Interoperability of Government Data Catalogues. EGOV 2010: 339-350.

at the Digital Enterprise Research Institute (DERI) and the Greek National Institute for Public Administration and Decentralization. DCAT is an RDF[8] 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) 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 section 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 methodology[9] 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 sections.

- In section 2, a number of related activities are identified.
- Section 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.
- Section 4 contains a reference to the base specification of the DCAT vocabulary, on which the Application Profile is based.
- In section 5, the terminology and the namespaces that are used in the specification of the Application Profile are introduced.

[8] W3C. Resource Description Framework (RDF). http://www.w3.org/RDF/
[9] European Commission. Joinup. Process and Methodology for Developing Core Vocabularies. http://joinup.ec.europa.eu/elibrary/document/isa-deliverable-process-and-methodology-developing-core-vocabularies

- The classes defined for the Application Profile are identified in section 6.
- Section 7 lists the mandatory, recommended and optional properties of those classes used in the Application Profile.
- In section 8, controlled vocabularies are proposed for use as value sets for a number of properties.
- Section 9 contains the Conformance Statement for this Application Profile.
- Accessibility and multilingual issues are addressed in section 10.
- In section 11, a number of pointers are given that may be helpful for implementation of the Application Profile in a Linked Data[10] environment.
- Finally, acknowledgements related to the development of this Application Profile are contained in section 12.

2. RELATED WORK

2.1. G8 Open Data Charter

In June 2013, G8 leaders signed the Open Data Charter[11]. The Open Data Charter sets out 5 strategic principles that all G8 members will act on. These include an expectation that all government data will be published openly by default, alongside principles to increase the quality, quantity and re-use of the data that is released. G8 members have also identified 14 high-value areas – from education to transport, and from health to crime and justice – from which they will release data. These will help unlock the economic potential of open data, support innovation and provide greater accountability.

While the DCAT Application Profile is intended as target for mapping across national metadata schemas in Europe, the G8 mappings[12] are intended to document how national and regional approaches around the world relate to a number of general metadata concepts. The G8 mappings consider both the terms displayed to the users of national and regional portals and the machine-readable terms used in the back-end of those portals. In the near future, a mapping of the DCAT Application Profile may be added to the G8 mappings.

2.2. Models for describing datasets or similar information

2.2.1. Asset Description Metadata Schema (ADMS)

The Asset Description Metadata Schema (ADMS)[13] is a vocabulary to describe interoperability assets (resources like specifications, schemas, code lists, software tools that facilitate interoperability) making it possible for ICT developers to

[10] W3C. Linked Data. http://www.w3.org/standards/semanticweb/data
[11] G8 Open Data Charter. https://www.gov.uk/government/publications/open-data-charter
[12] Github G8_Metadata_Mapping. https://github.com/nsinai/G8_Metadata_Mapping/blob/master/index.md
[13] European Commission. Joinup. Asset Description Metadata Schema (ADMS), version 1.00. 18 April 2012. https://joinup.ec.europa.eu/asset/adms/release/100

discover and re-use those assets. The ADMS namespace document[14] is published by W3C. The class of ADMS Asset is modelled as a subclass of DCAT Dataset.

2.2.2. CERIF for Datasets (C4D)

CERIF[15] is a European Union (EU) recommendation to Member States that defines a data model and XML interchange format for interoperability of research information, maintained by EuroCRIS[16] and used by more than 150 institutional systems across Europe and 10 national systems.

The overall aim of CERIF for Datasets (C4D)[17] 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.

2.2.3. CKAN Dataset Schema

The Comprehensive Knowledge Archive Network (CKAN)[18] is a Web-based open source data management system for the distribution of data maintained by the Open Knowledge Foundation[19]. The Dataset[20] is the central domain object in the CKAN Domain Model[21].

2.2.4. INSPIRE Metadata Schema

INSPIRE[22] is a Directive[23] 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 for metadata, data, network services, data and service sharing, monitoring and reporting. Such specifications consist of a set of implementing rules (which take the

[14] W3C. Asset Description Metadata Schema (ADMS). Namespace Document 25 June 2012. http://www.w3.org/ns/adms
[15] euroCRIS. CERIF Introduction. http://www.eurocris.org/Index.php?page=CERIFintroduction&t=1
[16] EuroCRIS Current Research Information Systems, the European Organisation for International Research Information. http://www.eurocris.org/
[17] JISC. CERIF for Datasets (C4D). Delivery scheduled 31 March 2013. http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/infrastructure/c4d.aspx
[18] CKAN. http://ckan.org/
[19] Open Knowledge Foundation. http://okfn.org/
[20] CKAN Dataset Schema. http://docs.ckan.org/en/ckan-1.8/domain-model-dataset.html
[21] CKAN Domain Model. http://docs.ckan.org/en/ckan-1.8/domain-model.html
[22] European Commission – Joint Research Centre. INSPIRE Web site: http://inspire.ec.europa.eu/
[23] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). OJ L 108, 25.4.2007, p. 1–14. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32007L0002:EN:NOT

form of Commission Regulations, i.e., they are legally binding in the EU Member States), along with the corresponding technical guidelines, defined by a regulatory committee composed of representatives of both EU Member States and European Union bodies and institutions.

The INSPIRE Metadata Implementing Rules include rules for the description of datasets.

2.2.5. Schema.org

Schema.org[24] is an activity that provides a collection of schemas, i.e., HTML tags, which 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 hierarchy[25] 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.

2.2.6. Statistical Data and Metadata eXchange (SDMX)

Statistical Data and Metadata eXchange (SDMX)[26] 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 17369:2013[27].

2.2.7. Vocabulary of Interlinked Datasets (VoID)

The Vocabulary of Interlinked Datasets (VoID)[28] is an RDF 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 Terms[29]), 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.

[24] Schema.org. http://schema.org/
[25] Schema.org. The Type Hierarchy. http://schema.org/docs/full.html
[26] Statistical Data and Metadata eXchange (SDMX). http://sdmx.org/
[27] ISO 17369:2013. Statistical data and metadata exchange (SDMX). http://www.iso.org/iso/catalogue_detail.htm?csnumber=52500
[28] W3C. Describing Linked Datasets with the VoID Vocabulary. http://www.w3.org/TR/void/
[29] Dublin Core Metadata Initiative. DCMI Metadata Terms. http://dublincore.org/documents/dcmi-terms/

2.3. Application Profiles

2.3.1. Austrian Open Government Data Cooperation

The Austrian Open Government Data Cooperation[30] has produced a DCAT Application Profile called "OGD Metadaten – 2.1"[31].

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

The "*Norma Técnica de Interoperabilidad de Reutilización de recursos de la información*"[32] 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).

2.3.3. OGD Metadata Structure of govdata.de

The OGD Metadata Structure[33] of govdata.de is a CKAN Profile. It is written as a JSON schema document. It has controlled vocabularies for topics and licences.

2.3.4. Project Open Data Common Core Metadata Schema

Project Open Data[34] 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 Schema[35] is provided defining 'required', 'required-if-applicable' and 'expanded' fields for the description of Datasets, based on DCAT.

2.3.5. WMO Core Metadata Profile

The WMO Core Metadata Profile[36] of the World Meteorological Organization[37] 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/Geomatics[38].

[30] Austrian Open Government Data Cooperation http://reference.e-government.gv.at/Open-Government-Data.2771.0.html
[31] OGD Metadaten – 2.1 http://reference.e-government.gv.at/Veroeffentlichte-Informationen.2774.0.html
[32] 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
[33] OGD Metadata Structure of govdata.de http://htmlpreview.github.io/?https://github.com/fraunhoferfokus/ogd-metadata/blob/master/OGPD_JSON_Schema.html
[34] Project Open Data. http://project-open-data.github.io/
[35] Project Open Data. Common Core Metadata Schema. http://project-open-data.github.io/schema/
[36] 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
[37] World Meteorological Organization. http://www.wmo.int/
[38] 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** datasets. This can be achieved by the exchange of descriptions of datasets 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 datasets 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, harmonisation, and publication services. The Open Data Support[39] project funded by the European Commission will operate a Metadata Broker service for data portals in Europe, it will use the DCAT Application Profile as a common metadata vocabulary

- **Data Consumers**: Users (data consumers) use the data portal of their choice to search through various collections of datasets from a single point of access. The data portal allows the user to explore (FRSAD – Functional Requirements for Subject Authority Records[40]), find, identify and select (FRBR – Functional Requirements for Bibliographic Records[41]) datasets coming from different EU Member States, different portals and different organisations. Data consumers could also be systems (machines).

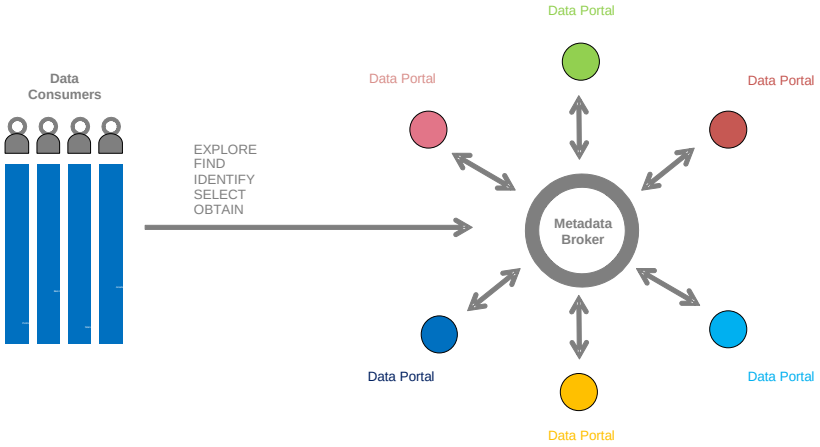

Figure 1 - Basic use case: enable a search for datasets across various data portals

[39] Open Data Support. https://joinup.ec.europa.eu/node/62928
[40] IFLA. Functional Requirements for Subject Authority Data (FRSAD). http://www.ifla.org/en/node/1297
[41] IFLA. Functional Requirements for Bibliographic Records. http://www.ifla.org/publications/functional-requirements-for-bibliographic-records

3.1. User Scenario 1 – Cross-data portal search for datasets on tourist accommodations

Tomasz works for a Polish tourist agency and is looking for datasets on tourist accommodations in European cities. His employer wants to develop a travel app to expand its service offerings. Unfortunately, the datasets 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 datasets about tourist accommodations. Unfortunately, data portals do not exchange their collections of datasets 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 datasets 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 datasets, 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.

This user scenario is enabled by the DCAT Application Profile, which plays the role of the common metadata vocabulary.

3.2. User Scenario 2 – Cross-data portal search for datasets on employment rates, immigration, and immigration control legislation

Julie works for a university and is looking for datasets 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 datasets and related legislation are made available by both national and EU 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 the DCAT Application Profile 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 EU 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 the DCAT Application Profile 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 the 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.

This user scenario is enabled by the DCAT Application Profile, which plays the role of the common metadata vocabulary.

3.3. User Scenario 3 – Cross-data portal search for datasets: cross-lingual aspects

The above-mentioned basic use case for cross-data portal search can be supplemented with various requirements for cross-lingual search. This is included in the following user scenario.

Bart wants to find all datasets in Belgium related to street names – maintained by local administrations – from a single point of access. As Belgium has 3 official languages (Dutch, French and German) –sometimes data owners include English descriptions as well –, the search for datasets should return all relevant datasets even when description metadata is only available in one of the aforementioned languages. To enable description metadata (categories, regions, ...) to be searchable in several languages the following requirements exist with regards to the representation and exchange of description metadata between data portal:

- **Multilingual descriptions**: a textual description is available in one or more languages;
- **Multilingual links**: link(s) to web pages with more information are available in one or more languages; and
- **Multilingual datasets**: the dataset itself is only available in one or more languages.

It is often helpful that a developer searching for a dataset gets all this information 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.

A brief account of how DCAT-AP deals with *representing* and *exchanging* multilingual description metadata is included in section 10. How cross-lingual search is handled in systems, for example in indexing and user interface presentation is outside of the scope of this Application Profile.

3.4. User Scenario 4 – Cross-data portal search for datasets: subscription and recommendation

The above-mentioned basic use case for cross-data portal search can be supplemented with features for content subscription and recommendation.

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.

Section 7 gives an overview of properties in DCAT-AP such as geographic coverage, theme and keyword that can be used to enter a subscription or make a recommendation. How subscriptions and recommendations are handled in systems is outside the scope of this Application Profile.

3.5. User Scenario 5 – Cross-data portal search for datasets: cross-border spatial queries for datasets

The above-mentioned basic use case for cross-data portal search can be refined with requirements for cross-border spatial search.

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 Geoportal[42] 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, licence, etc. of the datasets in the search results. The search results may contain a graphical representation of the spatial coverage on a map.

Section 7 gives an overview of properties in DCAT-AP such as geographic coverage and temporal coverage that can be used to filter search results.

[42] European Commission – Joint Research Centre. INSPIRE Geoportal. http://inspire-geoportal.ec.europa.eu/

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

The above-mentioned basic use case for cross-data portal search can be supplemented with requirements to retrieve contact information for a dataset.

Whenever someone accesses a dataset, it may happen that the user notes that there is an error in the dataset. In such cases, the user may want to contact the maintainer of the dataset to suggest correction of the error.

Section 7 describes the property adms:contactPoint, which allows to include contact information of the maintainer which the user can use to send an e-mail with an error notification.

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

This user scenario deals with requirements for exchanging description metadata between data portals.

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 web sites of aggregators. This mechanism can be based on the exchange of catalogue records (A) or on the exchange of an entire snapshot (B).

Mechanism A. Exchange based on catalogue records: A set of catalogue 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 catalogue records that represent the latest create, update, and delete transactions to its metadata.

2. **Exchange (push or pull)**: Periodically, the Data Portal pushes the catalogue 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.

3. **Update by Metadata Broker**: The Metadata Broker updates its own metadata to reflect the changes indicated in the catalogue records.
    - **Created records:** It will create the metadata for all catalogue 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 catalogue 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 catalogue records that indicate a deletion of metadata. For example, if a dataset 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 catalogue records as the Data Portal. In turn, it can offer a Catalogue Record-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 catalogue 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 catalogue 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 updates neither the metadata nor the corresponding catalogue 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 catalogue 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 catalogue record to reflect the creation.
    - **Updated metadata:** The Metadata Broker updates metadata which has been updated and also updates the modification date of the catalogue 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 catalogue 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 catalogue 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 catalogue records.
4. **Recordkeeping by Metadata Broker**: The Metadata Broker now can offer a catalogue record-based exchange of metadata.

Mechanism A is enabled by DCAT-AP by the use of the class dcat:CatalogRecord and its property *adms:status*, described Section 6 and 7 respectively. Mechanism B does not necessitate the use of the property adms:status.

3.8. User Scenario 8 – Cross-data portal search for datasets: 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.

- **Without DCAT Application Profile**: 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 available at: http://www.normattiva.it/static/mappa.html.
- **With DCAT Application Profile**: 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.

This user scenario could be enabled by the DCAT Application Profile, which plays the role of the common metadata vocabulary.

4. THE DCAT SPECIFICATION

The specification of the Data Catalog Vocabulary (DCAT) will be published by W3C as a W3C Recommendation. A Second Last Call Working Draft was published on 1 August 2013[43].

The DCAT Application Profile defined in this document is based on the version of 1 August 2013. No major changes are expected in the last call review at W3C.

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.
- **Recommended 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, if it is available.
- **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.

[43] W3C. Data Catalog Vocabulary (DCAT). Last Call Working Draft 01 August 2013. http://www.w3.org/TR/2013/WD-vocab-dcat-20130801/

- **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[44].

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.).

Classes are classified as 'Mandatory' in section 6.1 if they appear as the range of one of the mandatory properties in section 7.

The class 'Distribution' is classified as 'Recommended' in section 6.2 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.

All other classes are classified as 'Optional' in section 6.3. A further description of the optional classes is only included as a sub-section in section 7 if the Application Profile specifies mandatory or recommended properties for them.

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

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

6. APPLICATION PROFILE CLASSES

Figure 2 shows a UML diagram of all classes and properties included in the DCAT Application Profile.

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

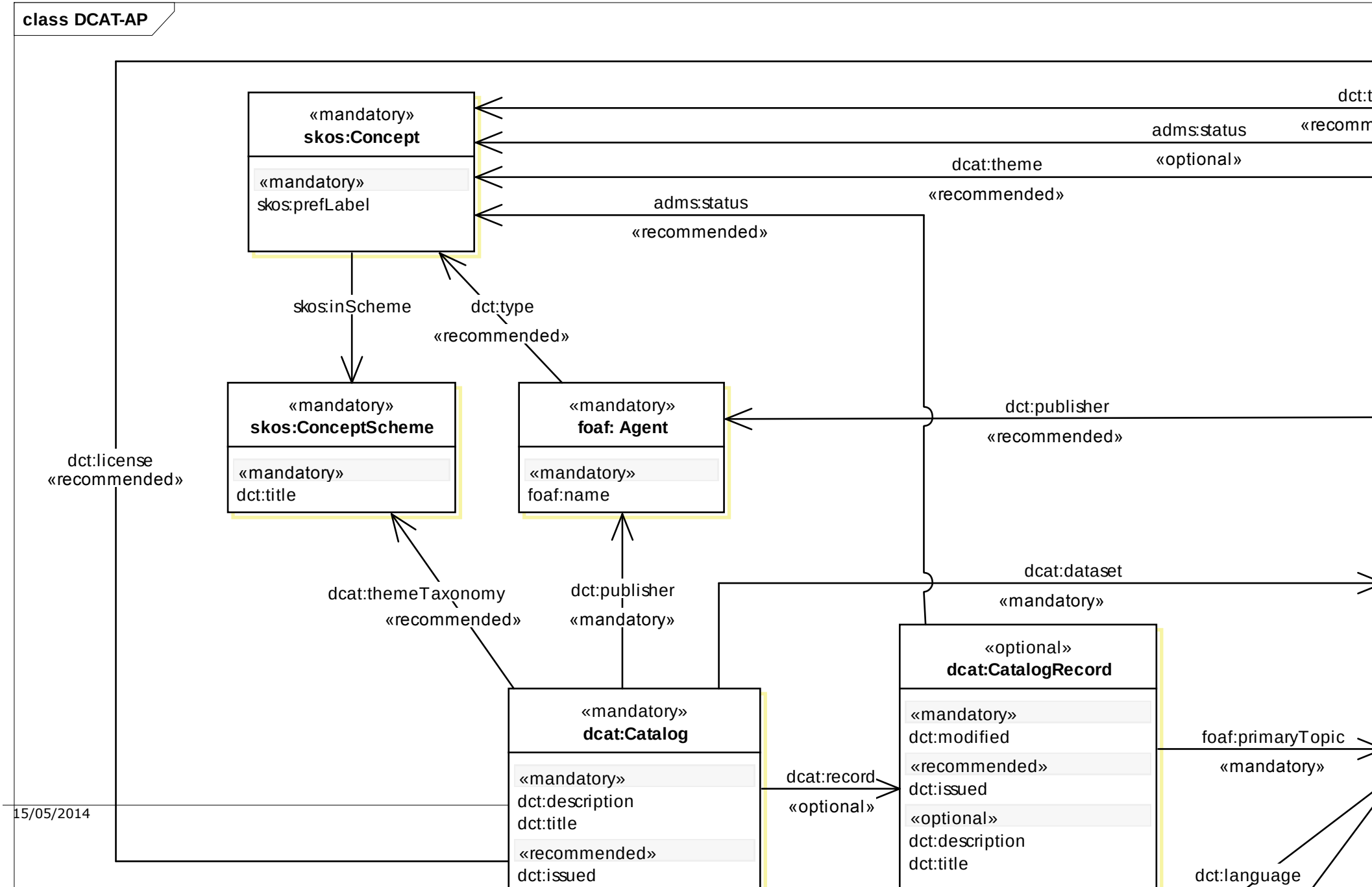
class DCAT-AP
«mandatory»
skos:Concept
«mandatory»
skos:prefLabel
«mandatory»
skos:ConceptScheme
«mandatory»
dct:title
«mandatory»
foaf: Agent
«mandatory»
foaf:name
«mandatory»
dcat:Catalog
«mandatory»
dct:description
dct:title
«recommended»
dct:issued
«optional»
dcat:CatalogRecord
«mandatory»
dct:modified
«recommended»
dct:issued
«optional»
dct:description
dct:title
skos:inScheme
dct:type
«recommended»
adms:status
«recommended»
dct:theme
«recommended»
adms:status
«optional»
dct:license
«recommended»
dct:publisher
«recommended»
dcat:themeTaxonomy
«recommended»
dct:publisher
«mandatory»
dcat:dataset
«mandatory»
dcat:record
«optional»
foaf:primaryTopic
«mandatory»
dct:language

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 Organization Ontology[45] 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
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[46].	rdfs:Literal	http://www.w3.org/TR/rdf-concepts/#section-Literals
Resource	Anything described by RDF.	rdfs:Resource	http://www.w3.org/TR/rdf-schema/#ch_resource

6.2. Recommended Class

Class name	Usage note for the Application Profile	URI	Reference
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

6.3. Optional Classes

Class name	Reason for exclusion	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
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
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

[45] W3C. The Organization Ontology. W3C Candidate Recommendation,25 June 2013. http://www.w3.org/TR/2013/CR-vocab-org-20130625/

[46] IETF. BCP 47. Tags for Identifying Languages. http://www.rfc-editor.org/rfc/bcp/bcp47.txt

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[47] is recommended.	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 conforms	dct:Standard	http://dublincore.org/documents/dcmi-terms/#terms-Standard
Status	An indication of the maturity of a Distribution.	skos:Concept	http://www.w3.org/TR/vocab-adms/#status
VCard	A description following the vCard specification, e.g. to provide telephone number and e-mail address for a contact point.	v:VCard	http://www.w3.org/2006/vcard/ns-2006.html#VCard

7. APPLICATION PROFILE PROPERTIES PER CLASS

A quick reference table of properties per class is included in Annex I.

7.1. Catalogue

7.1.1. Mandatory properties for Catalogue

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

7.1.2. 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

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

Property	URI	Range	Usage note	Card.
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 (KOS) 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 changed or modified.	0..1

7.1.3. Optional properties for Catalogue

Property	URI	Range	Usage note	Card.
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

7.2. Catalogue Record

7.2.1. 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

7.2.2. Recommended properties for Catalogue Record

Property	URI	Range	Usage note	Card.
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
change type	adms:status	skos:Concept	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

7.2.3. 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
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
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
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

7.3.3. Optional properties for Dataset

Property	URI	Range	Usage note	Card.
conforms to	dct:conformsTo	dct:Standard	This property refers to an implementing rule or other specification.	0..n
frequency	dct:accrualPeriodicity	dct:Frequency	This property refers to the frequency at which Dataset is updated.	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 Catalogue.	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	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[48], DataCite[49], DOI[50], EZID[51] or W3ID[52].	0..n
release date	dct:issued	rdfs:Literal typed as xsd:dateTime	This property contains the date of formal issuance (e.g., publication) of the Dataset.	0..1

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

Property	URI	Range	Usage note	Card.
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
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	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 at the access URL may contain information about how to get the Dataset.	1..n

7.4.2. 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

7.4.3. 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
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
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
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

Property	URI	Range	Usage note	Card.
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

7.5. Agent

7.5.1. Mandatory property 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

7.5.2. Recommended property for Agent

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

7.6. Category Scheme

7.6.1. Mandatory property 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 property 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. Licence Document

7.8.1. Recommended property for Licence Document

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

7.9. Period Of Time

7.9.1. 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. 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)				

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 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.

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	Vocabulary name	Vocabulary URI	Usage note
dcat:mediaType	Distribution	MDR File types Name Authority List[53]	http://publications.europa.eu/mdr/authority/file-type/	
dcat:theme	Dataset	EuroVoc domains[54]	http://eurovoc.europa.eu/100142 through 100162	
dcat:themeTaxonomy	Catalogue	EuroVoc[55]	http://eurovoc.europa.eu/	
dct:accrualPeriodicity	Dataset	Dublin Core Collection Description Frequency Vocabulary[56]	http://purl.org/cld/freq/	
dct:format	Distribution	MDR File Type Named Authority List	http://publications.europa.eu/mdr/authority/file-type/	
dct:language	Catalogue, Dataset	MDR Languages Named Authority List[57]	http://publications.europa.eu/mdr/authority/language/	
dct:publisher	Catalogue, Dataset	MDR Corporate bodies Named Authority List[58]	http://publications.europa.eu/mdr/authority/corporate-body/	To 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 Countries Named Authority List[59], MDR Places Named Authority List[60]	http://publications.europa.eu/mdr/authority/country/, http://publications.europa.eu/mdr/authority/place/	The Countries vocabulary is to be used if the scope is a particular country. The Places vocabulary is to be used if the scope is a part of a country.
adms:status	Catalogue Record	ADMS change type vocabulary	http://purl.org/adms/changetype/	:created, :updated, :deleted
adms:status	Distribution	ADMS status vocabulary	http://purl.org/adms/status/	The list of terms in the ADMS status vocabulary is included in the ADMS specification[61]
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

[53] Publications Office of the European Union. Metadata Registry. Authorities. File types. http://publications.europa.eu/mdr/authority/file-type/
[54] European Union. EuroVoc domains and microthesauri. http://eurovoc.europa.eu/drupal/?q=node/555
[55] European Union. EuroVoc, the European Union's multilingual thesaurus. http://eurovoc.europa.eu/drupal/
[56] Dublin Core Metadata Initiative. Dublin Core Collection Description Frequency Vocabulary. http://dublincore.org/groups/collections/frequency/
[57] Publications Office of the European Union. Metadata Registry. Authorities. Languages. http://publications.europa.eu/mdr/authority/language/
[58] Publications Office of the European Union. Metadata Registry. Authorities. Corporate bodies. http://publications.europa.eu/mdr/authority/corporate-body/
[59] Publications Office of the European Union. Metadata Registry. Authorities. Countries. http://publications.europa.eu/mdr/authority/country/
[60] Publications Office of the European Union. Metadata Registry. Authorities. Places. http://publications.europa.eu/mdr/authority/place/
[61] European Commission. Joinup. Asset Description Metadata Schema (ADMS). ADMS 1.00. http://joinup.ec.europa.eu/asset/adms/release/100

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 vocabularies[62], the Dewey Decimal Classification[63] and numerous other schemes.

8.4. Licence vocabularies

Concerning licence vocabularies, implementers are encouraged to use widely recognised licences such as Creative Commons licences[64], and in particular the CC Zero Public Domain Dedication[65], the Open Data Commons Public Domain Dedication and License (PDDL)[66], the ISA Open Metadata Licence[67], the European Union Public Licence (EUPL)[68] or an open government licence such as the UK Open Government Licence[69].

Further activities in this area are undertaken by the Open Data Institute[70] with the Open Data Rights Statement Vocabulary[71] and by the Open Digital Rights Language (ODRL) Initiative[72].

9. CONFORMANCE STATEMENT

9.1. Provider requirements

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

- Provide a description of the Catalogue, including at least the mandatory properties specified 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.

[62] http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.5/CERIF1.5_Semantics.xhtml
[63] OCLC. Dewey Summaries as Linked Data. http://www.oclc.org/dewey/webservices.en.html and http://dewey.info/
[64] Creative Commons. About The Licenses. http://creativecommons.org/licenses/
[65] Creative Commons. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. http://creativecommons.org/publicdomain/zero/1.0/
[66] Open Data Commons Public Domain Dedication and License (PDDL). http://opendatacommons.org/licenses/pddl/
[67] ISA Open Metadata Licence v1.1, https://joinup.ec.europa.eu/category/licence/isa-open-metadata-licence-v11
[68] European Commission. Joinup. Open Source Software. European Union Public Licence (EUPL). http://joinup.ec.europa.eu/software/page/eupl
[69] The National Archives. Open Government Licence for public sector information. http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/
[70] Open Data Institute. http://www.theodi.org/
[71] Open Data Institute. Open Data Rights Statement Vocabulary. http://schema.theodi.org/odrs/
[72] Open Digital Rights Language (ODRL) Initiative. http://www.w3.org/community/odrl/

- Provide descriptions of Distributions, if any, 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.

For the properties listed in the table in section 8, 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 in section 7 MAY be provided.

9.2. Receiver requirements

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

- Process information for all classes specified in section 6.
- Process information for all properties specified in section 7.
- Process information for all controlled vocabularies specified in section 8.2.

As stated in section 6, "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.).

10. Accessibility and Multilingual Aspects

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

Multilingual aspects related to this Application Profile concern all properties whose contents are expressed as strings 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, 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.

The requirement in section 3.3 for multilingual links can be met through a content negotiation[73] mechanism whereby different content is served based on the Accept-Languages indicated by the browser. Using such a mechanism, the link to the landing page can resolve to different language versions of the web page with more information about the Dataset.

How multilingual information is handled in systems, for example in indexing and user interfaces, 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 Data[74] environment, publishers should consider the recommendations in the W3C Notes "Best Practice Recipes for Publishing RDF Vocabularies"[75], "Best Practices for Publishing Linked Data"[76] and the ISA report "10 Rules for Persistent URIs"[77].

Publishers should also consider it to be best practice to assign URIs to all instances of the classes described in section 6.

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.

[73] Apache Web Server: content negotiation. http://httpd.apache.org/docs/current/content-negotiation.html
[74] W3C. Linked Data. http://www.w3.org/standards/semanticweb/data
[75] W3C. Best Practice Recipes for Publishing RDF Vocabularies. http://www.w3.org/TR/swbp-vocab-pub/
[76] W3C. Best Practices for Publishing Linked Data. https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html
[77] European Commission. Joinup. 10 Rules for Persistent URIs. https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris

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.
- 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 SPARQL[78], OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting)[79], the Atom Publishing Protocol[80], the Atom Syndication Format[81] with 'tombstones'[82], the Protocol for the Syndication of Resource Descriptions (SDShare)[83], the Data Catalog Interoperability Protocol[84] 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 datasets have been created, updated, or deleted from the data provider's catalogue.

This Application Profile does not consider requirements for tracking provenance of metadata or data, other than providing information about the publisher of the data. If additional provenance information is required, implementers are encouraged to consider the use of W3C PROV Ontology[85] to capture and exchange such information. For instance, it can be used to model roles different from the data publisher, like "creator", "processor", "maintainer", "rights holder", and "user".

[78] W3C. SPARQL Query Language for RDF. http://www.w3.org/TR/rdf-sparql-query/
[79] Open Archives Initiative. Protocol for Metadata Harvesting. http://www.openarchives.org/pmh/
[80] IETF. RFC 5023. The Atom Publishing Protocol. http://www.ietf.org/rfc/rfc5023.txt
[81] IETF. RFC 4287. The Atom Syndication Format. http://www.ietf.org/rfc/rfc4287.txt
[82] IETF. RFC 6721. The Atom "deleted-entry" Element. http://www.ietf.org/rfc/rfc6721.txt
[83] SDShare. A Protocol for the Syndication of Resource Descriptions. http://www.sdshare.org/spec/sdshare-20120710.html
[84] Data Catalog Interoperability Protocol. http://spec.datacatalogs.org/
[85] W3C. PROV-O: The PROV Ontology. http://www.w3.org/TR/prov-o/

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:

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 Research	date.berlin.de	DE
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
De Keyzer	Michiel	PwC EU Services	Open Data Support	BE
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
García García	Emilio	Ministry of Finance and Public Administrations	Datos.gob.es	ES
Goedertier	Stijn	PwC EU Services	Open Data Support	BE
Gutteridge	Christopher	University of Southhampton	data.southampton.ac.uk	UK
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
Stefano	Nativi	Institute for Atmospheric Pollution, Italian National Research Council (IIA-CNR)	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
Prieto	Eva	Red.es	Datos.gob.es	ES
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
Thoeye	Thimo	City of Ghent - E-Strategy	Open Data Gent	BE
Van der Waal	Sander	Open Knowledge Foundation	publicdata.eu	UK

The members of the Review Group:

Family name	First name	Organisation	Portal	
Hanssens	Bart	Fedict	Data.gov.be	BE
Kolsjö	Magnus	PwC		SE
Quertinmont	Jean-Charles	Agence pour la Simplification Administrative	psi.belgium.be	BE
Sicila	Miguel-Ángel	Eurocris	eurocris.org	EU

ANNEX I. QUICK REFERENCE OF CLASSES AND PROPERTIES

Class	Class URI	Mandatory properties	Recommended properties	Optional properties
Agent	foaf:Agent	foaf:name	dct:type	
Category	skos:Concept	skos:prefLabel		
Category Scheme	skos:ConceptScheme	dct:title		
Catalogue	dcat:Catalog	dcat:dataset dct:description dct:publisher dct:title	dct:issued dct:language dct:license dct:modified dcat:themeTaxonomy foaf:homepage	dcat:record dct:rights dct:spatial
Catalogue Record	dcat:CatalogRecord	dct:modified foaf:primaryTopic	adms:status dct:issued	dct:description dct:title
Dataset	dcat:Dataset	dct:description dct:title	adms:contactPoint dcat:distribution dcat:keyword dcat:theme dct:publisher	adms:identifier adms:version adms:versionNotes dcat:landingPage dct:accrualPeriodicity dct:conformsTo dct:identifier dct:issued dct:language dct:modified dct:spatial dct:temporal
Distribution	dcat:Distribution	dcat:accessURL	dct:description dct:format dct:license	adms:status dcat:byteSize dcat:downloadURL dcat:mediaType dct:issued dct:modified dct:rights dct:title
Document	foaf:Document			
Frequency	dct:Frequency			
Identifier	adms:Identifier			
Licence Document	dct:LicenseDocument	dct:type		
Licence Type	skos:Concept			
Linguistic System	dct:LinguisticSystem			
Literal	rdfs:Literal			
Location	dct:Location			
Media Type or Extent	dct:MediaTypeOrExtent			
Period Of Time	dct:PeriodOfTime		schema:startDate schema:endDate	
Publisher Type	skos:Concept			
Resource	rdfs:Resource			
Rights Statement	dct:RightsStatement			
Standard	dct:Standard			
Status	skos:Concept			
VCard	v:VCard			