# DCAT Application Profile for data portals in Europe

## Revision Draft 1

Date: 23/06/2015

# Document Metadata

| Property | Value |
|---|---|
| Date | 2015-04-02 |
| Status | Revision Draft 1 |
| Version | 0.07 |
| Rights | © 2015 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 |

# 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.

This document is the result of a revision in 2015 of the first version of the agreement that was published in September 2013[5].

## 1.2. Scope of the revision

This objective of this work is to produce and updated release of the DCAT Application Profile based on requests for change coming from the real-world implementation of the specification.

The Application Profile specified in this document is based on the specification of the **Data Catalog Vocabulary** (DCAT)[6] developed under the responsibility of the Government Linked Data Working Group[7] at W3C. The work on DCAT was initiated[8] at the Digital Enterprise Research Institute (DERI) and the Greek National Centre for Public Administration and Decentralization. DCAT is an RDF[9] 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.

[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] https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-application-profile-data-portals-europe-final

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

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

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

[9] W3C. Resource Description Framework (RDF). http://www.w3.org/RDF/

The Working Group that worked on the revision of the Application Profile took into account the following criteria in deciding which changes to make to the specification:

(1) **Improve the discovery of datasets.** Priority was given to those aspects that can be used in searching and navigating and requests that had to do with dataset management and operational issues were not considered. Some implementation advice of a general nature was included but no further deployment advice for specific operational environments was added.

(2) **Ensure compliance with DCAT.** Requests that aimed at changing the model of re-model DCAT or replacing properties that already existed in DCAT by others with similar semantics were not taken into account. Only elements have been added that could be re-used from existing namespaces and no new terms were created in a new namespace.

(3) **Ensure simplicity: keep a narrow scope.** Only elements were added for which there was evidence that they were needed and that the information was commonly available in existing systems; if there were several ways to meet a requirement, the simplest one was chosen.

(4) **Ensure application domain neutrality.** Requests that were only relevant for certain types of datasets were not considered. In particular, requirements that were related to geographic characteristics and to handling of statistical data were referred to parallel activities, GeoDCAT and StatDCAT.

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 7.

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. Change management process

The work is conducted according to the change management process for structural metadata that was developed by the ISA Programme of the European Commission (ISA Programme - European Commission, 2015). Firstly, the methodology prescribes a governance structure for managing changes and releases. Secondly, it specifies specific processes for managing changes in specifications, for managing the preparation of releases of specification, and for managing the process of publication of a releases.

Based on the governance structure as prescribed in the ISA methodology, the following governance levels, activities and roles have been identified for the revision process of the DCAT Application Profile:

**Table 1: Governance structure for the revision of DCAT-AP**

| Governance level | Activities | Who |
|---|---|---|
| **Steering Committee (SC)** | ▪ Approve general directions<br>▪ Stay informed about activities and progress<br>▪ Endorse new release | ISA Coordination Group, PSI Expert Group (DG CNECT) |

| | | |
|---|---|---|
| **Governance Committee (GC)** | ▪ Organise & safeguard proper execution of maintenance activities<br>▪ Identify the need for revision<br>▪ Instruct the Operational Team to apply changes<br>▪ Establish Working Groups<br>▪ Prepare release for Steering Committee endorsement | ISA Programme Management Team |
| **Operational Team (OT)** | ▪ Gather change requests<br>▪ Advise Governance Committee on nature of changes<br>▪ Provide the editor for Working Groups<br>▪ Document the resolution of change requests | Contractor of ISA action 1.1. |
| **Working Group (WG)** | ▪ Bring expertise<br>▪ Reach consensus | - Chair: Norbert Hohn, Willem Van Gemert (Publications Office)<br>- Editor: Makx Dekkers<br>- Organisations implementing the specification<br>- Individual experts |

The ISA change management methodology identified five phases in the change management and release process. These phases will apply to the revision of the DCAT Application Profile:

1. **Request handling**
   This phase starts with the receipt of requests for change (RFC) from stakeholders. A log of all change requests received will be made available via Joinup[10]. The requests are evaluated by the Operational Team (OT) and grouped into issues on Joinup[11]. Based on the analysis by the OT, the Governance Committee (GC) decides on the further process. If the request is rejected because it is not clear or not relevant for the specification at hand, the GC informs the submitter of the rejection with a justification. If the request is accepted, the GC will schedule the request for inclusion in a new release.

2. **Request resolution**
   In order to resolve the requests for semantic changes, the GC establishes a Working Group (WG). The WG elaborates one or more drafts of the revised specification and discusses these drafts until consensus is reached. It then submits the draft to the GC who publishes the draft for public review. The WG resolves any comments and finalises the new specification. The process continues with the Release preparation phase.

3. **Release preparation**
   The GC instructs the OT to prepare the specification and any additional documentation. The GC notifies the Steering Committee (SC) that the new release is ready for publication and requests endorsement by the SC.

[10] Request analysis: https://joinup.ec.europa.eu/sites/default/files/dcat-ap_revision_requests_v0.03.xlsx
[11] DCAT-AP Issue Log: https://joinup.ec.europa.eu/asset/dcat_application_profile/issue/all

4. **Release endorsement**
   The SC discusses the new release and endorses its publication.

5. **Release publication**
   Following endorsement by the SC, the GC publishes the new release and notifies the stakeholders and the wider public of its availability. The new release of the DCAT-AP will be made available on Joinup[12].

### 1.4. The DCAT specification

The specification of the Data Catalog Vocabulary (DCAT) was published by W3C as a W3C Recommendation on 16 January 2014[13].

The DCAT Application Profile defined in this document is based on this officially published version.

[12] Revised DCAT-AP on Joinup: https://joinup.ec.europa.eu/node/137964
[13] W3C. Data Catalog Vocabulary (DCAT). W3C Recommendation 16 January 2014. http://www.w3.org/TR/2014/REC-vocab-dcat-20140116/

## 2. TERMINOLOGY USED IN THE APPLICATION PROFILE

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.

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.
- **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[14].

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 3.1 if they appear as the range of one of the mandatory properties in section 5.

The class 'Distribution' is classified as 'Recommended' in section 3.2 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.

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

All other classes are classified as 'Optional' in section 3.3. A further description of the optional classes is only included as a sub-section in section 5 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/
- owl: http://www.w3.org/2002/07/owl#
- rdfs: http://www.w3.org/2000/01/rdf-schema#
- schema: http://schema.org/
- skos: http://www.w3.org/2004/02/skos/core#
- spdx: http://spdx.org/rdf/terms#
- xsd: http://www.w3.org/2001/XMLSchema#
- vcard: http://www.w3.org/2006/vcard/ns#

Application Profile classes Figure 1 shows a UML diagram of all classes and properties included in the DCAT Application Profile.

New diagram to be created.

Figure 1 - DCAT Application Profile UML Class Diagram

## 3. APPLICATION PROFILE CLASSES

### 3.1. 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[15] 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[16]. | 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 |

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

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

## 3.2. Recommended Classes

| 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 |
| 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 |

## 3.3. 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 |
| Checksum | A value that allows the contents of a file to be authenticated. This class allows the results of a variety of checksum and cryptographic message digest algorithms to be represented. | spdx:Checksum | http://spdx.org/rdf/terms#Checksum |
| Document | A textual resource intended for human consumption that contains information, e.g. a web page about a Dataset. | foaf:Document | http://xmlns.com/foaf/spec/#term_Document |
| Frequency | A rate at which something recurs, e.g. the publication of a Dataset. | dct:Frequency | http://dublincore.org/documents/dcmi-terms/#terms-Frequency |
| Identifier | An identifier in a particular context, consisting of the string that is the identifier; an optional identifier for the identifier scheme; an optional identifier for the version of the identifier scheme; an optional identifier for the agency that manages the identifier scheme | adms:Identifier | http://www.w3.org/TR/vocab-adms/#identifier |
| 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[17] 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 or the type of change of a CatalogRecord. | skos:Concept | http://www.w3.org/TR/vocab-adms/#status |

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

| | | | |
|---|---|---|---|
| Kind | A description following the vCard specification, e.g. to provide telephone number and e-mail address for a contact point. Note that the class Kind is the parent class for the four explicit types of vCards (Individual, Organization, Location, Group). | vcard:Kind | http://www.w3.org/TR/2014/NOTE-vcard-rdf-20140522/#d4e181 |

## 4. CHANGE LOG

The table below summarises the changes applied to the current release of the DCAT-AP

| URI | Type | Action | Description | Issue |
|---|---|---|---|---|
| adms:sample | Optional property (Dataset) | New | Range: rdfs:Resource<br>Cardinality: 0..n<br>This property refers to a sample of the data. | PR21 |
| adms:versionNotes | Optional property (Dataset) | Updated | *Updates*<br>Cardinality: 0..1 → 0..n<br>This property can be repeated for parallel language versions of the version notes. | CO4 |
| dcat:contactPoint | Recommended property (Dataset) | Updated | *Updates*<br>URI: adms:contactPoint → dcat:contactPoint<br>Range: VCard → vCard:Kind | MO8 |
| dct:creator | Optional property (Dataset) | New | Range: foaf:Agent<br>This property refers to the authority under whose responsibility the Dataset is made available | CO5 |
| dct:license | Optional property (CatalogRecord) | New | Range: dct:LicenseDocument<br>This property refers to the licence under which the metadata for the Dataset is made available. | (Lee) |
| dct:LicenseDocument | Recommended class | Updated | *Updates*<br>Optional class → recommended class | OB4 |
| dct:source | Optional property (CatalogRecord) | New | Range: dcat:CatalogRecord<br>This property links to the original metadata that was used in creating metadata for the Dataset | MO5 |
| dct:valid | Optional property (Dataset) | New | Range: dct:PeriodOfTime<br>This property refers to a temporal period for which the Dataset is valid. | PR18 |
| owl:versionInfo | Optional property (Dataset) | Updated | *Updates*<br>URI: adms:version → owl:versionInfo | CO4 |
| spdx:algorithm | Mandatory property (Checksum) | New | Range: spdx:checksumAlgorithm_sha1<br>Cardinality: 1..1<br>This property identifies the algorithm used to produce the subject Checksum. Currently, SHA-1 is the only supported algorithm. It is anticipated that other algorithms will be supported at a later time. | PR1 |
| spdx:Checksum | Optional class | New | A value that allows the contents of a file to be authenticated. This class allows the results of a variety of checksum and cryptographic message digest algorithms to be represented. | PR1 |
| spdx:checksum | Optional property (Dataset) | New | Range: spdx:Checksum<br>Cardinality: 0..1<br>This property provides a mechanism that can be used to verify that the contents of a Distribution have not changed | PR1 |
| spdx:checksumValue | Mandatory property (Checksum) | New | Range: hexBinary<br>Cardinality: 1..1<br>This property provides a lower case hexadecimal encoded digest value produced using a specific algorithm. | PR1 |
| vCard:Kind | Optional class | Updated | *Updates*<br>Class name: VCard → Kind<br>URI: VCard → vcard:Kind<br>The class Kind is the parent class for the four explicit types of vCards (Individual, Organization, Location, Group). | MO4 |

# 5. DCAT Application Profile Properties per Class

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

## 5.1. Catalogue

### 5.1.1. 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 8. | 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 |

### 5.1.2. 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 |
| 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 |

### 5.1.3. 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 |

## 5.2. Catalogue Record

### 5.2.1. 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 |

| 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 Catalogue entry was changed or modified. | 1..1 |

#### *5.2.2. 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 |

#### *5.2.3. 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 |
| licence | dct:license | dct:LicenseDocument | This property refers to the licence under which the metadata for the Dataset is made available. | 0..1 |
| source metadata | dct:source | dcat:CatalogRecord | This property links to the original metadata that was used in creating metadata for the Dataset | 0..1 |
| title | dct:title | rdfs:Literal | This property contains a name given to the Catalogue. This property can be repeated for parallel language versions of the name. | 0..n |

### 5.3. Dataset

#### *5.3.1. 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 |

#### *5.3.2. Recommended properties for Dataset*

| Property | URI | Range | Usage note | Card |
|---|---|---|---|---|
| contact point | dcat:contactPoint | vcard:Kind | This property contains contact information that can be used for 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 |

### 5.3.3. 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 |
| authority | dct:creator | foaf:Agent | This property refers to the authority under whose responsibility the Dataset is made available | 0..1 |
| 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[18], DataCite[19], DOI[20], EZID[21] or W3ID[22]. | 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 |
| 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 |
| valid | dct:valid | dct:PeriodOfTime | This property refers to a temporal period for which the Dataset is valid. | 0..n |
| version | owl:versionInfo | rdfs:Literal | This property contains a version number or other version designation of the Dataset. | 0..1 |
| version notes | adms:versionNotes | rdfs:Literal | This property contains a description of the differences between this version and a previous version of the Dataset. This property can be repeated for parallel language versions of the version notes. | 0..n |

## 5.4. Distribution

### 5.4.1. 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 |

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

### 5.4.2. 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 |

### 5.4.3. 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 |
| checksum | spdx:checksum | spdx:Checksum | This property provides a mechanism that can be used to verify that the contents of a distribution have not changed | 0..1 |
| 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 |
| sample | adms:sample | rdfs:Resource | This property refers to a sample of the data | 0..n |
| status | adms:status | skos:Concept | This property refers to the maturity of the Distribution | 0..1 |
| title | dct:title | rdfs:Literal | This property contains a name given to the Distribution. This property can be repeated for parallel language versions of the description. | 0..n |
| update/ modification date | dct:modified | rdfs:Literal typed as xsd:date or xsd:dateTime | This property contains the most recent date on which the Distribution was changed or modified. | 0..1 |

## 5.5. Agent

### 5.5.1. 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 |

### 5.5.2. Recommended property for Agent

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

## 5.6. Category Scheme

### 5.6.1. 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 |

## 5.7. Category

### *5.7.1. 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 |

## 5.8. Checksum

### *5.8.1. Mandatory properties for Checksum*

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

## 5.9. Licence Document

### *5.9.1. 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 |

## 5.10. Period Of Time

### *5.10.1. 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.<br>The start of the period should be understood as the start of the date, hour, minute etc. given (e.g. starting at midnight at the beginning of the day if the value is a date); the end of the period should be understood as the end of the date, hour, minute etc. given (e.g. ending at midnight at the end of the day if the value is a date) | | | | |

# 6. CONTROLLED VOCABULARIES

## 6.1. Requirements for controlled vocabularies

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

Controlled vocabularies SHOULD:

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

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

### 6.2. 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[23] | http://publications.europa.eu/mdr/authority/file-type/ | |
| dcat:theme | Dataset | EuroVoc domains[24] Under discussion | http://eurovoc.europa.eu/100142 through 100162 | |
| dcat:themeTaxonomy | Catalogue | EuroVoc[25] Under discussion | http://eurovoc.europa.eu/ | |
| dct:accrualPeriodicity | Dataset | Dublin Core Collection Description Frequency Vocabulary[26] Under discussion | http://purl.org/cld/freq/ | |
| dct:format | Distribution | MDR File Type Named Authority List Under discussion | http://publications.europa.eu/mdr/authority/file-type/ | |

[23] Publications Office of the European Union. Metadata Registry. Authorities. File types. http://publications.europa.eu/mdr/authority/file-type/
[24] European Union. EuroVoc domains and micro-thesauri. http://eurovoc.europa.eu/drupal/?q=node/555
[25] European Union. EuroVoc, the European Union's multilingual thesaurus. http://eurovoc.europa.eu/drupal/
[26] Dublin Core Metadata Initiative. Dublin Core Collection Description Frequency Vocabulary. http://dublincore.org/groups/collections/frequency/

| Property URI | Used for Class | Vocabulary name | Vocabulary URI | Usage note |
|---|---|---|---|---|
| dct:language | Catalogue, Dataset | MDR Languages Named Authority List[27] | http://publications.europa.eu/mdr/authority/language/ | |
| dct:publisher | Catalogue, Dataset | MDR Corporate bodies Named Authority List[28] Under discussion | 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[29], MDR Places Named Authority List[30] | 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[31] |
| 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 |

### 6.3. Other controlled vocabularies

In addition to the proposed common vocabularies in section 6.2, 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[32], the Dewey Decimal Classification[33] and numerous other schemes.

### 6.4. Licence vocabularies

Concerning licence vocabularies, implementers are encouraged to use widely recognised licences such as Creative Commons licences[34], and in particular the CC Zero Public Domain Dedication[35], the Open Data Commons Public Domain Dedication

[27] Publications Office of the European Union. Metadata Registry. Authorities. Languages. http://publications.europa.eu/mdr/authority/language/
[28] Publications Office of the European Union. Metadata Registry. Authorities. Corporate bodies. http://publications.europa.eu/mdr/authority/corporate-body/
[29] Publications Office of the European Union. Metadata Registry. Authorities. Countries. http://publications.europa.eu/mdr/authority/country/
[30] Publications Office of the European Union. Metadata Registry. Authorities. Places. http://publications.europa.eu/mdr/authority/place/
[31] European Commission. Joinup. Asset Description Metadata Schema (ADMS). ADMS 1.00. http://joinup.ec.europa.eu/asset/adms/release/100
[32] http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.5/CERIF1.5_Semantics.xhtml
[33] OCLC. Dewey Summaries as Linked Data. http://www.oclc.org/dewey/webservices.en.html and http://dewey.info/
[34] Creative Commons. About The Licenses. http://creativecommons.org/licenses/
[35] Creative Commons. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. http://creativecommons.org/publicdomain/zero/1.0/

and License (PDDL)[36], the ISA Open Metadata Licence[37], the European Union Public Licence (EUPL)[38] or an open government licence such as the UK Open Government Licence[39].

Further activities in this area are undertaken by the Open Data Institute[40] with the Open Data Rights Statement Vocabulary[41] and by the Open Digital Rights Language (ODRL) Initiative[42].

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

# 7. CONFORMANCE STATEMENT

## 7.1. Provider requirements

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

- Provide a description of the Catalogue, including at least the mandatory properties specified in section 5.1.1.
- Provide information for the mandatory properties specified in section 5.2.1, 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 5.3.1.
- Provide descriptions of Distributions, if any, of Datasets in the Catalogue, including at least the mandatory properties specified in section 5.4.1.
- Provide descriptions of all organisations involved in the descriptions of Catalogue and Datasets, including at least the mandatory properties specified in section 5.5.1.
- 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 5.6.1.
- Provide descriptions of all categories involved in the descriptions of Datasets in the Catalogue, including at least the mandatory properties specified in section 5.7.1.

For the properties listed in the table in section 6, 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 5 MAY be provided.

## 7.2. Receiver requirements

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

- Process information for all classes specified in section 3.
- Process information for all properties specified in section 5.
- Process information for all controlled vocabularies specified in section 6.2.

As stated in section 3, "processing" means that receivers must accept incoming data and transparently provide these data to applications and services. It does neither imply nor prescribe what applications and services finally do with the data (parse, convert, store, make searchable, display to users, etc.).

# 8. 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[43] 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.

[43] Apache Web Server: content negotiation. http://httpd.apache.org/docs/current/content-negotiation.html

## 9. ACKNOWLEDGEMENTS

This work was elaborated by a Working Group under the ISA programme. The Working Group was chaired by **Norbert Hohn** and **Willem Van Gemert** from the Publications Office. The ISA Programme of the European Commission was represented by **Vassilios Peristeras** and **Athanasios Karalopoulos**. **Makx Dekkers** was the editor of the specification.

The members of the Working Group:

| Family name | First name | Organisation | Portal | |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

## 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<br>dct:description<br>dct:publisher<br>dct:title | dct:issued<br>dct:language<br>dct:license<br>dct:modified<br>dcat:themeTaxonomy<br>foaf:homepage | dcat:record<br>dct:rights<br>dct:spatial |
| Catalogue Record | dcat:CatalogRecord | dct:modified<br>foaf:primaryTopic | adms:status<br>dct:issued | dct:description<br>dct:license<br>dct:soorce<br>dct:title |
| Checksum | spdx:Checksum | spdx:algorithm<br>spdx:checksumValue | | |
| Dataset | dcat:Dataset | dct:description<br>dct:title | dcat:contactPoint<br>dcat:distribution<br>dcat:keyword<br>dcat:theme<br>dct:publisher | adms:identifier<br>adms:versionNotes<br>dcat:landingPage<br>dct:accrualPeriodicity<br>dct:conformsTo<br>dct:creator<br>dct:identifier<br>dct:issued<br>dct:language<br>dct:modified<br>dct:spatial<br>dct:temporal<br>owl:versionInfo |
| Distribution | dcat:Distribution | dcat:accessURL | dct:description<br>dct:format<br>dct:license | adms:sample<br>adms:status<br>dcat:byteSize<br>dcat:downloadURL<br>dcat:mediaType<br>dct:issued<br>dct:modified<br>dct:rights<br>dct:title<br>spdx:checksum |
| Document | foaf:Document | | | |
| Frequency | dct:Frequency | | | |
| Identifier | adms:Identifier | | | |
| Kind | vcard:Kind | | hasEmail | hasTelephone |
| 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<br>schema:endDate | |
| Publisher Type | skos:Concept | | | |
| Resource | rdfs:Resource | | | |
| Rights Statement | dct:RightsStatement | | | |
| Standard | dct:Standard | | | |
| Status | skos:Concept | | | |