MS7. Compliance with the Metadata Quality Assurance (MQA) tool for datasets
Version 0.3

Document Information

Action Number	2018-EU-IA-0104
Project Website	https://saintgeorgeonabike.eu
Contractual Deadline	31/08/2020
Milestone	MS7
Document type	Report
Author	Nuno Freire, Antoine Isaac (EF)
Contributor(s)	Albin Larsson (EF)
Reviewer	Ariadna Lobo (BSC)
Keywords	Dataset Metadata, European Data Portal

Change Log

Version	Author	Description of Change
V0.1	Nuno Freire, Antoine Isaac (EF)	First iteration
V0.2	Ariadna Lobo (BSC)	Reviewed
V0.3 (ongoing)	Antoine Isaac (EF), Albin Larsson (EF)	Handling post-submission comments, updating specification following tests with EDP

Table of Contents

1. Introduction

Europeana has the role of facilitating the usage of Cultural Heritage (CH) resources from and about Europe. It currently holds metadata from over 4,000 providers.

Saint George on a Bike (SGoaB) aims to improve the quality and quantity of metadata associated with imagery from European cultural heritage, using Artificial Intelligence technology, with an emphasis on the collections held by Europeana. SGoaB's activity 4, "Adoption of DSI interoperability standards for Linked Open Data deployment", focuses on ensuring the interoperability of SGoaB's results (both services and data) with existing DSI (Digital Service infrastructure) and Work Programme requirements.

A specific focus is given to the European Data Portal (EDP)[1], whose mission is to collect the metadata of public sector information available on public data portals across European countries. The EDP addresses the whole data value chain: from data publishing to data re-use. This milestone document seeks to address the interoperability between SGoaB and the EDP.

At this stage of the project, however, use cases (MS3) and their implementation (MS5) are not finalized. Especially, there needs to be final decisions on which data enrichments the project is going to produce (data models, formats...) and how these are going to be bundled with Europeana's "regular" metadata and services for CH resources.

For this milestone document (MS7. Compliance with the Metadata Quality Assurance (MQA) tool for datasets), we consider a baseline case where SGoaB activities would result in the publication of enrichments to the Europeana metadata directly in the metadata published by Europeana. We therefore address here how to make the Europeana dataset available in the EDP.

This document first presents a high-level overview of the requirements for the EDP, then reminds some relevant facts about Europeana's context. Section 4 then dives in the detail of how to provide metadata on the Europeana dataset to the EDP, using the DCAT-AP standard. Section 5 presents some steps for validating the resulting metadata using EDP services. Section 6 presents the next steps of our work in SGoaB activity 4, which will be reflected in future iterations of this document.

2. Requirements for the European Data Portal

The European Data Portal provides a Metadata Quality Assurance (MQA) dashboard [2] to check the quality of the metadata submitted. We ultimately plan to submit all metadata submitted by SGoaB to this dashboard. However, the MQA is only a final step in the provision of data to EDP. This document therefore devotes much attention to guiding the preceding steps, especially how to adhere to the expected standards for dataset description and data harvesting.

EDP offers several technical alternatives for harvesting metadata from its providers.

---

[1] https://www.europeandataportal.eu/

[2] https://www.europeandataportal.eu/mqa/

All alternatives are described in EDP's public documentation[3], which presents the technical requirements and constraints of EDP's harvesting process. In the case of an implementation by Europeana, some alternatives are clearly less demanding since these are based on technologies currently in use in Europeana. Especially, given Europeana's in-house knowledge of OAI-PMH and RDF file dumps, our analysis is focused on them.

The EDP harvests metadata about (catalogues of) datasets, which can follow several alternative specifications. We have opted for providing the expected dataset metadata following the the DCAT Application Profile for data portals in Europe Version 2.0.0 (DCAT-AP)[4]. DCAT-AP is a specification for describing public sector datasets in Europe, and it is based on the Data Catalogue vocabulary (DCAT). The datasets' metadata is provided to EDP as linked data represented in RDF.

Under the metadata model in-use at EDP, Europeana must describe itself as a data catalogue hosting the metadata about the datasets that it aggregates. The metadata of the datasets should also describe how the datasets can be obtained - the dataset's distributions. The EDP does not harvest the dataset distributions[5], instead, the distributions are accessed by data re-users that discover the existence of the datasets via the EDP.

3. Europeana's (Meta)datasets

The Europeana dataset consists mostly of descriptive metadata about CH objects, but it also contains metadata about other kinds of entities used for contextualizing the CH objects.

This aggregated dataset is subdivided into disjoint sub-datasets. Each of these datasets is harvested by Europeana directly from CH institutions or from intermediary data aggregators.[6] A dataset may contain data that originates from only one cultural heritage institution, or from several ones (this is sometimes the case when the datasets are harvested from an aggregator).

Europeana's solution for data modelling is the Europeana Data Model (EDM). EDM supports several of the core processes of the Europeana network of data providers and aggregators. EDM also contributes to the access layer of the Europeana platform, supporting the sharing of data with third parties and allowing Europeana to be 'a big aggregation of digital representations of culture artefacts together with rich contextualization data and embedded in a linked Open Data architecture'. EDM has been developed (and is still being improved) in collaboration with the wider cultural heritage community, involving representatives from all the domains represented in

---

[3]

https://www.europeandataportal.eu/sites/default/files/edp_s1_gdl_data-supplier-guidelines_v9 .pdf accessible via https://www.europeandataportal.eu/en/about/be-harvested-us

[4] Currently in version 2.0.1, see DCAT Application Profile for data portals in Europe
[5] in the case of Europeana and SGoaB, these distributions would be the (enriched) metadata about the digitized CH objects accessible via Europeana.
[6] For a definition of these terms, please see the Europeana glossary of terms at https://pro.europeana.eu/page/glossary

Europeana, i.e. libraries, museums, archives, and galleries. EDM is based on the principles of the Web of Data [9] and can be represented in RDF.

EDM allows the descriptions of several classes of entities, including of course that of cultural objects. But for this document the most relevant class is that of datasets, which is described in the EDM Dataset profile[7].

4. Providing Europeana Datasets Metadata as DCAT-AP Metadata

DCAT-AP specifies mandatory, recommended and optional data elements. It is mandatory to provide an instance of a `dcat:Catalogue`, which must be accompanied with an instance of a `foaf:Agent` that represents the publisher of the catalogue. The dataset(s) that comprise the catalogue must be represented as instance(s) of `dcat:Dataset`. Additional classes may also be used when providing the data in these three core classes.

This section presents the classes of DCAT-AP used for mapping the metadata about Europeana's datasets. The mapping is made from the datasets metadata that Europeana maintains according to the EDM Dataset profile, plus additional data that should be generated for the specific purpose of contributing to the EDP.

The mappings are presented organized by the classes specified in DCAT-AP that Europeana should use. The mappings of each class are shown in tables that contain the following elements:

- *Mapping note*. Describes the purpose of the usage of the class and remarks about its mapping.
- *URI*. Describes how URI(s) for instances of the class should be assigned or created.
- *Properties*. The properties applicable to instances of the class, and information about the ranges for their values, their cardinality and specific notes about their mappings.

We indicate points where a decision is still needed by using To Be Confirmed (TBC) annotations.

DCAT-AP and EDM reuse terms from various existing specifications. For readability of the document, the classes and properties are prefixed accordingto their namespaces, as follows:

- adms: http://www.w3.org/ns/adms
- dcat: http://www.w3.org/ns/dcat#
- dcatap: http://data.europa.eu/r5r/
- dct: http://purl.org/dc/terms/
- edm: http://www.europeana.eu/schemas/edm/
- foaf: http://xmlns.com/foaf/0.1/
- rdfs: http://www.w3.org/2000/01/rdf-schema#
- skos: http://www.w3.org/2004/02/skos/core#
- xsd: http://www.w3.org/2001/XMLSchema#

---

[7] https://pro.europeana.eu/page/edm-profiles#organization-dataset-profiles

Note on metadata provision

This document does not yet address how the metadata on Europeana datasets will be created and provided to the EDP. This could be done via creating a file or set of files, manually or automatically, depending on the scope (especially, how many datasets are to be represented in the metadata). Neither do we address the issue of practical data transfer, i.e., whether the metadata will be handed in to EDP manually by Europeana, via a simple file sharing, or more automatically, via an (updatable) description that could be served by Europeana's APIs.

Class dcat:Catalogue

Class dcat:Catalogue			
Mapping note:	This class is used for the root resource of Europeana's catalogue of datasets.		
URI:	`http://data.europeana.eu/catalog#catalog`		
Property	**Range**	**Card.**	**Mapping note**
dcat:dataset	dcat:Dataset	1..n	TBC: refer to the aggregated dataset of Europeana and also for each of the individual datasets in Europeana OR to the aggregated dataset of Europeana.
dct:description	rdfs:Literal	1..n	The literal: `"The catalogue of all cultural heritage datasets aggregated by Europeana"@en`
dct:publisher	foaf:Agent	1..n	The value must be an instance of a `foaf:Agent` describing the Europeana Foundation.See the mapping for `foaf:Agent`.
dct:title	rdfs:Literal	1..1	The literal `"Europeana Datasets"@en`
foaf:homepage	foaf:Document	0..1	The URI `https://www.europeana.eu/`
dct:license	dct:LicenseDocument	0..1	The value should be the URI for CC0, `http://creativecommons.org/public domain/zero/1.0/` described separately as a resource of type `dct:LicenseDocument`. See the mapping for `dct:LicenseDocument`.
dct:issued	rdfs:Literal typed as xsd:date or xsd:dateTime	0..1	Must be a `xsd:date` or a `xsd:dateTime`. Automatically generated.
dct:language	dct:LinguisticSystem	0..n	The Europeana catalogue may contain data in multiple languages, depending on the languages in use in the descriptions for the datasets, therefore no value should be provided for dct:language.
dct:spatial	dct:Location	0..n	The URI of the entity for Europe defined in the EU Vocabularies: `http://publications.europa.eu/res`

			`ource/authority/continent/EUROPE`. See also the mapping for dct:Location.
dcat:themeTax onomy	skos:ConceptSc heme	0..n	The URI of the Dataset Theme Vocabulary: `http://publications.europa.eu/res ource/authority/data-theme`
dct:modified	rdfs:Literal typed as xsd:date or xsd:dateTime	0..1	Must be a `xsd:date` or a `xsd:dateTime`. Automatically generated
dct:creator	foaf:Agent	0..n	The value must be the instance of a `foaf:Agent` describing the Europeana Foundation. See the mapping for `foaf:Agent`.

Class dcat:Dataset

From the point of view of the EDP, Europeana can be seen as a catalogue of individual datasets from European CH institutions but for operational reasons (and should it prove not harmful to the visibility of the data in EDP) we could submit all Europeana data as one overarching, single dataset - the dataset aggregated by Europeana.

Instances of `dcat:Dataset` must be provided by Europeana in these two cases. The following two tables present the mappings for the two cases separately.

Class dcat:Dataset (for the dataset aggregated by Europeana)			
Mapping note:	An instance of this class is used to represent the Europeana aggregated dataset of cultural heritage digital objects. The class is mapped from `edm:Collection` defined in the Europeana Datasets Profile		
URI:	TBC: for the moment it seems that the European Data Portal can accept a blank node, which seems acceptable as long as Europeana does not publish its individual datasets.		
Property	**Range**	**Card.**	**Mapping note**
dct:description	rdfs:Literal	1..n	The value: `"A dataset containing metadata descriptions of cultural heritage digital objects from, and about, Europe. The dataset aggregates metadata from more than 4,000 cultural heritage institutions."@en` The property may be repeated for providing translations. The values should have a `xml:lang` tag.
dct:title	rdfs:Literal	1..n	The value: `"Europeana aggregated dataset of cultural heritage digital objects"@en` The property may be repeated for translations of the title. The values should

			have a `xml:lang` tag
dcat:distribution	dcat:Distribution	0..n	One value should be created specifying the distribution via the SPARQL API of Europeana. See the mapping for `dcat:Distribution`.
dct:publisher	foaf:Agent	0..1	The value should be the instance of `foaf:Agent` representing the Europeana Foundation. See the mapping for `foaf:Agent`.
dct:creator	foaf:Agent	0..n	The value should be the instance of `foaf:Agent` representing the Europeana Foundation. See the mapping for `foaf:Agent`.
dct:spatial	dct:Location	0..n	The URI of the entity for Europe defined in the EU Vocabularies: `http://publications.europa.eu/resource/authority/continent/EUROPE`. See also the section for dct:Location.
dct:conformsTo	dct:Standard	0..n	A value for the EDM schema, stated with the URI `http://www.europeana.eu/schemas/edm/`.
dcat:theme	skos:Concept	0..n	The URI for the category 'Education, Culture and Sport" of EDP's controlled vocabulary for dataset categories: `http://publicdata.eu/def/category/education-culture-and-sport` The list of EDP categories can be consulted at the end of this section.

Class dcat:Dataset (for individual datasets from CH institutions)	
Mapping note:	This class is used to represent the datasets in Europeana. The class is mapped from `edm:Collection` defined in the Europeana Datasets Profile.
URI:	TBC: The URIs should be assigned according to a pattern based on the Europeana "dataset name"

Property	Range	Card.	Mapping note
dct:description	rdfs:Literal	1..n	Mapped from the `dct:description`. The property may be repeated for translations of the description. The values should have a `xml:lang` tag.
dct:title	rdfs:Literal	1..n	Mapped from `edm:datasetName` The `edm:datasetName` does not provide a real human-readable title, however, These are strings like `"2022362_Ag_UK_CultureGrid_Royal_Museums_Greenwich"`, which cannot be

			considered as belonging to any language (and thus cannot be provided with an `xml:lang` tag). If we want to submit descriptions for Europeana individual (sub)datasets, however, there is no alternative but using it since the property is mandatory in DCAT-AP. TBC: if these values are not deemed acceptable by EDP, we will have to forego the submission of individual dataset metadata.
dcat:distribution	dcat:Distribution	0..n	One value should be created specifying the file dump created by Europeana for distributing the dataset. See the mapping for the `dcat:Distribution` class.
dct:publisher	foaf:Agent	0..1	The value should be the instance of `foaf:Agent` representing the Europeana Foundation.
dct:creator	foaf:Agent	0..n	Mapped from the `edm:dataProvider` and represented as a `foaf:Agent`. See mapping for the `foaf:Agent` class. TBC: in order to recognize contributions from Europeana's aggregators we also plan to create `dct:creator` statements for the `edm:provider` and `edm:intermediateProvider` organizations. `dct:creator` is however not a great fit for their role. If DCAT-AP would use other roles, like `dct:contributor`, then we could switch to using it. We could also use `dct:publisher` as a better (but not perfect) fit, if DCAT-AP would allow to list more than one per dataset.
dct:spatial	dct:Location	0..n	The EDM Datasets profile features an `edm:country` property that could be an interesting candidate for `dct:spatial`. But in effect it is used to indicate the country of a dataset's provider and therefore is a rather poor fit for the spatial coverage of a dataset.
dct:conformsTo	dct:Standard	0..n	A value for the EDM schema, stated with the URI `http://www.europeana.eu/schemas/edm/`.
dcat:theme	skos:Concept	0..n	One value should be provided using EDP's controlled vocabulary for dataset categories. Consult the table at the end of this section. TBC: Since the Datasets Profile does not contain any property for categorizing the dataset, all datasets should have the most generally applicable category in Europeana value for dcat:theme:

			`http://publicdata.eu/def/category` `/education-culture-and-sport`. There might still be ways to automatically assign categories, but this requires more investigation.

The European Data Portal uses a controlled vocabulary for dataset categories. which are stated by dcat:theme properties. Below, are the categories used and their corresponding URIs:

Category	URI
Agriculture, fisheries, forestry and food	http://publicdata.eu/def/category/agriculture-fisheries-forestry-and-food
Education, culture and sport	http://publicdata.eu/def/category/education-culture-and-sport
Environment	http://publicdata.eu/def/category/environment
Energy	http://publicdata.eu/def/category/energy
Transport	http://publicdata.eu/def/category/transport
Science and technology	http://publicdata.eu/def/category/science-and-technology
Economy and finance	http://publicdata.eu/def/category/economy-and-finance
Population and society	http://publicdata.eu/def/category/population-and-society
Health	http://publicdata.eu/def/category/health
Government and public sector	http://publicdata.eu/def/category/government-and-public-sector
Regions and cities	http://publicdata.eu/def/category/regions-and-cities
Justice legal system and public safety	http://publicdata.eu/def/category/justice-legal-system-and-public-safety

TBC: in the future, we may want to add references to Europeana's usage guidelines[8] and attribution[9] pages.

Class dcat:Distribution

Instances of `dcat:Distribution` should be provided by Europeana in two cases. First to represent the SPARQL endpoint of Europeana, and second, to represent the file-based distributions of the individual datasets from CH institutions. The tables below present the mappings for the two cases separately.

Class dcat:Distribution (for the distributions for the whole Europeana dataset via SPARQL endpoint, OAI-PMH API and data dumps)

---

[8] https://www.europeana.eu/en/rights/usage-guidelines-for-metadata
[9] https://www.europeana.eu/en/rights/europeana-data-sources

Mapping note:	These instances of `dcat:Distribution` are used to represent the access to the complete aggregated dataset of Europeana via Europeana's SPARQL endpoint, OAI-PMH and data dumps.
URI:	For the time being these instances are going to be treated as unnamed individuals (blank nodes) TBC: maybe revisit this in a future version

Property	Range	Card.	Mapping note
dcat:accessService	dcat:DataService	0..n	The URI of the `dcat:DataService` representing the SPARQL endpoint, OAI-PMH API and data dumps of Europeana. See the mapping for `dcat:DataService`.
dcat:accessURL	rdfs:Resource	1..n	The distribution specific access URI such as `https://pro.europeana.eu/page/sparql`
dct:description	rdfs:Literal	0..n	The value "`Data access via the Europeana <DISTRIBUTION NAME>`"@en
dct:license	dct:LicenseDocument	0..1	The value should be the URI for CC0, `http://creativecommons.org/publicdomain/zero/1.0/` described separately as a resource of type `dct:LicenseDocument`. See the mapping for `dct:LicenseDocument`.
dct:conformsTo	dct:Standard	0..n	Two values should be provided for dct:conformsTo. A value for the EDM schema, stated with the URI `http://www.europeana.eu/schemas/edm/`. Another value for distribution specific standards. For example, SPARQL with the URI `http://www.w3.org/TR/sparql11-query/`.
adms:status	skos:Concept	0..1	The value must be an URI defined by ADMS such as `http://purl.org/adms/status/UnderDevelopment` or `http://purl.org/adms/status/Completed`
dcatap:availability	skos:Concept	0..1	The value must be the URI defined by DCAT-AP: `http://data.europa.eu/r5r/stable`.

Class dcat:Distribution (for each individual datasets of Europeana)	
Mapping note:	This class is used to represent the file dumps and other relevant distributions that Europeana publishes for each of its datasets.

URI:	TBC		
Property	**Range**	**Card.**	**Mapping note**
dcat:accessURL	rdfs:Resource	1..n	The URI `ftp://download.europeana.eu/dataset`
dct:description	rdfs:Literal	0..n	The value: `"A ZIP archive containing the dataset in EDM"@en`
dct:license	dct:LicenseDocument	0..1	The value should be the URI for CC0, `http://creativecommons.org/publicdomain/zero/1.0/` described separately as a resource of type `dct:LicenseDocument`. See the mapping for `dct:LicenseDocument`.
dcat:byteSize	rdfs:Literal typed as xsd:decimal	0..1	Automatically generated.
spdx:checksum	spdx:Checksum	0..1	Automatically generated.
dcat:downloadURL	rdfs:Resource	0..n	The value is created from the template: `ftp://download.europeana.eu/dataset/XML/<dataset_id>.zip`
dct:conformsTo	dct:Standard	0..n	A value for the EDM schema, stated with the URI `http://www.europeana.eu/schemas/edm/`.
dcat:mediaType	dcs:MediaTypeOrExtent	0..1	The value: `application/rdf+xml`
dcat:packageFormat	dcs:MediaTypeOrExtent	0..1	The value: `application/zip`
adms:status	skos:Concept	0..1	For all datasets published by Europeana the value must be the URI defined by ADMS: `http://purl.org/adms/status/Completed`
dct:modified	rdfs:Literal typed as xsd:date or xsd:dateTime	0..1	Automatically generated.
dcatap:availability	skos:Concept	0..1	The value must be the URI defined by DCAT-AP: `http://data.europa.eu/r5r/available`. (as the Europeana datasets can be updated then the data may be accessible from another distribution)

Class dcat:DataService

Several instances of `dcat:DataService` may be provided by Europeana to describe the APIs that provide public access to its data distributions. This section presents the mapping for the dcat:DataService instance that represents the SPARQL endpoint of Europeana.

Class dcat:DataService (for the Europeana SPARQL API and the OAI-PMH API)			
Mapping note:	This class is used to represent the Europeana SPARQL API.		
URI:	TBC: For the time being instances of `dcat:DataService` are going to be treated as unnamed individuals (blank nodes) because each service corresponds to one dataset.		
Property	**Range**	**Card.**	**Mapping note**
dct:description	rdfs:Literal	1..n	One of the literals: "`The SPARQL API of Europeana APIs allows exploring connections between Europeana data and outside data sources, like VIAF, Iconclass, Getty Vocabularies (AAT), Geonames, Wikidata, and DBPedia. It allows one to delve into the structured metadata of Europeana (for instance, to ask the question \"What are all the French 18th-century painters with at least five artworks available through Europeana\")"`@en, "`The Europeana OAI-PMH service offers a way to collect large amounts of Europeana data from our repository through a protocol named OAI-PMH which stands for Open Archives Initiative Protocol for Metadata Harvesting. The OAI-PMH protocol (presently in v2.0) allows you to harvest the entirety, or a selection, of all Europeana metadata so that it can be integrated into other services or applications.`"@en
dct:title	rdfs:Literal	1..n	One of the literals: "`Europeana SPARQL API`"@en, "`Europeana OAI-PMH API`"
dcat:endpointURL	rdfs:Resource	0..n	The value must be the URI: `http://sparql.europeana.eu/` or `https://api.europeana.eu/oai/record`

dcat:servesDataset	dcat:Dataset	0..n	TBC: we are going to populate this property when URIs for datasets are created (which is not the case right now)

Class foaf:Agent

Instances of `foaf:Agent` must be provided by Europeana in two cases. First to represent the Europeana Foundation, and second, to represent the CH institutions, when Europeana would provide description for individual (provider-level) datasets. The tables below present the mappings for the two cases separately.

Class foaf:Agent (for the Europeana Foundation)			
Mapping note:	This class is used to represent the Europeana Foundation.		
URI:	`http://data.europeana.eu/organization/1482250000001617026`		
Property	**Range**	**Card.**	**Mapping note**
foaf:name	rdfs:Literal	1..n	The property should have two literal values: `"Europeana Foundation"@en` and `"Stichting Europeana"@nl` It may be repeated for translations of the name. The values should have a `xml:lang` tag.
dct:type	skos:Concept	0..1	The value should be the URI for non-profit organization defined by ADMS: `http://purl.org/adms/publishertype/NonProfitOrganisation`

Class foaf:Agent (for the cultural heritage institutions)			
Mapping note:	This class is used to represent the cultural heritage institutions. These instances of `foaf:Agent` are used in properties of individual Europeana datasets. TBC: Instances of `foaf:Agent` should also be created for providers and intermediate providers.		
URI:	TBC: Europeana is in the process of minting and publishing identifiers for its data providers (via the Entity API), which we will use here when individual datasets are described.		
Property	**Range**	**Card.**	**Mapping note**
foaf:name	rdfs:Literal	1..n	The value corresponds to either the property `edm:dataProvider`, `edm:provider` or `edm:intermediateProvider` that give rise to the creation of the Agent resource. The property may be repeated for translations of the name. The values should have a `xml:lang` tag.

TBC: once we can attach organisation data to DCAT-AP records, we will revisit this mapping.

Class dct:LicenseDocument

Class dct:LicenseDocument			
Mapping note:	This class is used to represent the licenses of the europeana catalogue and of the datasets' distributions it lists. It is CC0, therefore the type of the resource shall be 'public domain' in the classification suggested by DCAT-AP (see below)		
URI:	Since all data distributed by Europeana is made available under CC0 the URI must be 'http://creativecommons.org/publicdomain/zero/1.0/".		
Property	**Range**	**Card.**	**Mapping note**
dct:type	skos:Concept	0..n	The URI of the public domain license define by ADMS: `http://purl.org/adms/licencetype/PublicDomain`

Class dct:Location

The `dct:Location` class is the range defined by DCAT-AP for dct:spatial properties of catalogues and datasets. However, a fully fledged representation of individual `dct:Location` resources is not necessary since DCAT-AP requires that the EU Vocabularies' Name Authority Lists are used for continents, countries and places that are in those lists, and this is the case for the locations we foresee in our metadata. If a particular location is not in one of the mentioned Named Authority Lists, Geonames URIs must be used.

Class dct:Standard

The `dct:Standard` class is the range defined by DCAT-AP for dct:conformsTo properties of datasets and distributions. DCAT-AP does not specify any properties to be used for instances of dct:Standard, therefore we only focus here on identifying the URIs to refer to the standards in use by Europeana.

One of the uses of `dct:conformsTo` is to express the usage of the Europeana Data Model, or EDM, as the schema of the datasets. In this case, the URI of the EDM namespace must be used: `http://www.europeana.eu/schemas/edm/`.

Another use of `dct:conformsTo` is to express the conformity of the distributions of the Europeana datasets via the SPARQL endpoint, the OAI-PMH API or data dumps. In this case, for example, the URI of the SPARQL query specification must be used: `http://www.w3.org/TR/sparql11-query/`.

5. Validation of the Europeana catalogue metadata

This section is based on a message sent to us by Gianfranco Cecconi (European Data Portal) on testing dataset descriptions. We copy it here with minor adaptations, as it is very fit for describing what we will do wrt. validation.

The metadata of the Europeana dataset catalogue should be validated on the validator service provided by the EDP. This validation is performed by submitting the DCAT-AP metadata, via an HTTP POST request, to the OpenAPI SHACL validator[10]. The service responds with the SHACL validation report for the submitted DCAT-AP metadata.

For invoking the OpenAPI, a tool such as Postman (https://www.getpostman.com/) can be used in the following way:

1. Create a new "request"

---
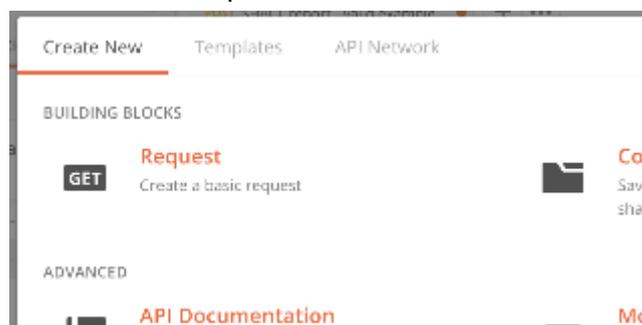
[10] Additional technical details for using the OpenAPI SHACL validator are available at https://www.europeandataportal.eu/shacl/

2. Give the request a name and put it in a collection
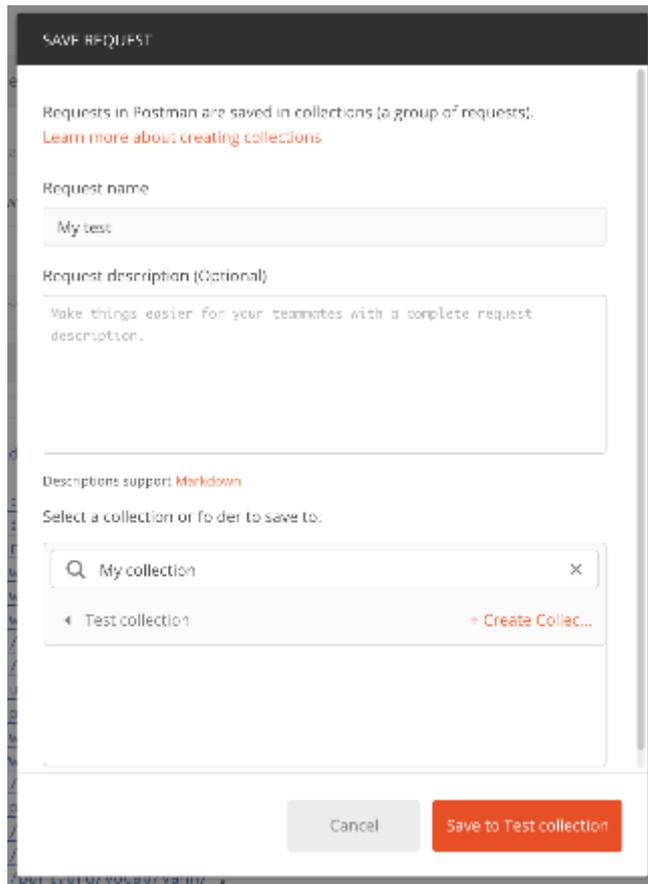
3. Choose to use HTTP POST and use as the request URL:
   https://www.europeandataportal.eu/shacl/validation/report/
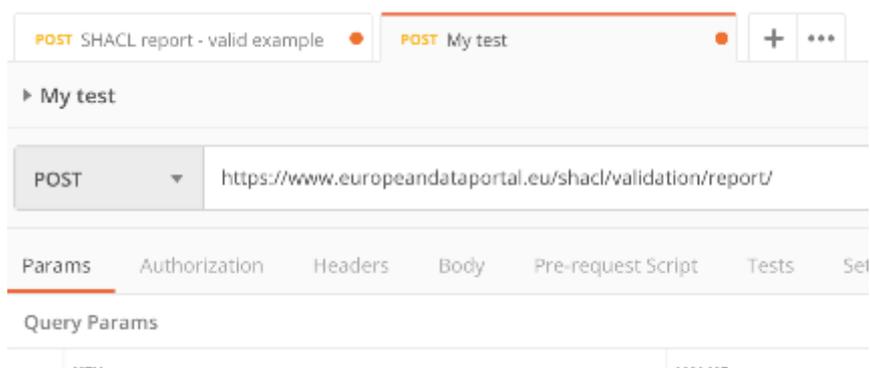
4. Specify the "Content-Type" variables in the HTTP headers according to the type of DCAT-AP representation in use, typically "application/rdf+xml"
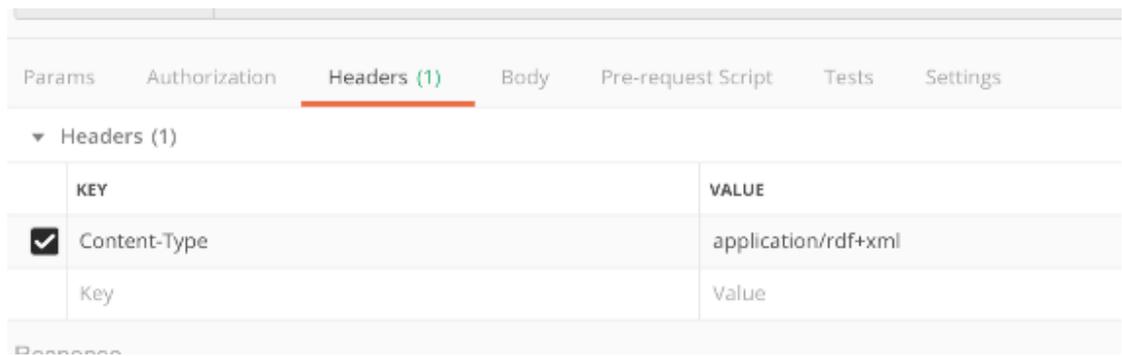
5. Copy and paste the DCAT-AP RDF in the "body" to be sent, ensuring to specify that the "raw" input is being used:
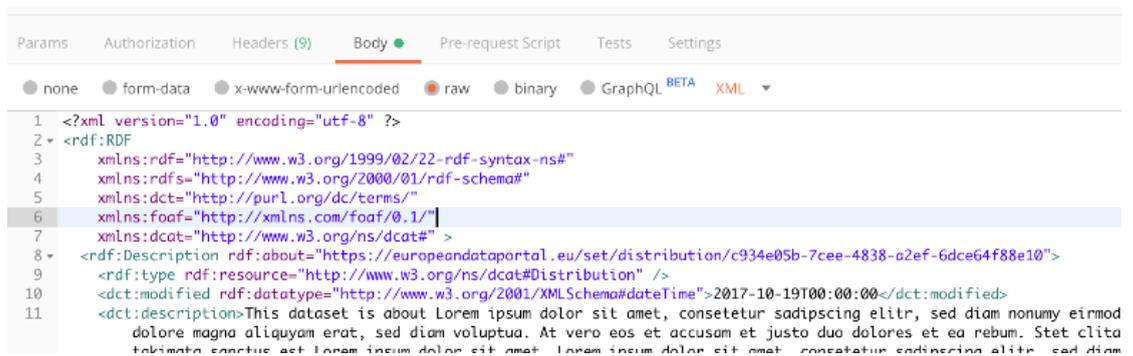
6. Click on the "Send" button
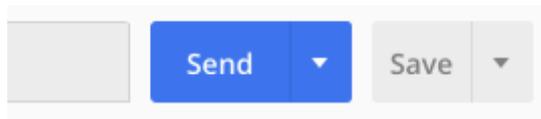
7. Examine the results in the bottom pane. If the DCAT-AP is valid, the result will be an empty report in JSON-LD that looks like:

Otherwise, if the metadata contains validation errors, they will be listed. Every error has an entry in the JSON-LD file that looks like the following:

8. Then, amend the DCAT-AP to address the errors and go back to step 5 until all validation errors are solved.

6. Implementation status

Europeana has created a draft of the DCAT-AP representation outlined by this

document (Appendix A). In its current state the DCAT-AP representation only describes the Europeana dataset as a whole and not each dataset separately.

This draft passes the OpenAPI SHACL validator in Section 5 without SHACL violations but with three SHACL warnings.

7. Next steps

For the future work of SGoaB activity 4 and the future iterations of this document, we plan a number of actions already penciled in earlier sections:

- To fine tune the mapping and make decisions over pending (TBC) issues
- To define the URIs to be used within the DCAT-AP representation, which are not yet minted by Europeana
- To finalize the DCAT-AP representation as per section 4 and finally undergo the tests of the Metadata Quality Assurance dashboard.
- Begin to address the handling of SGoaB enrichments potentially separately from Europeana's original metadata, depending on the decision of how these enrichments are handled in the architecture (see MS5 document). This could possibly be done through uploading the SGoaB data to a third-party platform supporting the output of DCAT-AP (like Zenodo[11]).

In addition, we will investigate whether some of Europeana's data is already present in the EDP, in case it has been submitted via another route, and how to handle it. This can happen when a Europeana data provider or aggregator also contributes to a national data portal that contributes to the EDP. Two known cases would be data from the national CH aggregators DDB[12] and SOCH[13] that contribute to the general national aggregators of their respective countries.

---

[11] https://about.zenodo.org/principles/

[12] https://www.europeandataportal.eu/data/datasets/b47f7ffa-9246-4a2a-9663-bf2cf9889883?locale=en

[13] https://www.europeandataportal.eu/data/datasets?query=&locale=en&keywords=museum&keywords=kultur&keywords=historia&minScoring=0&page=1

Appendix A. Draft of a DCAT-AP representation for Europeana data

```
@prefix dcat:   <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix foaf:   <http://xmlns.com/foaf/0.1/> .
@prefix vcard:  <http://www.w3.org/2006/vcard/ns#> .
@prefix adms:   <http://www.w3.org/ns/adms> .
@prefix dcatap: <http://data.europa.eu/r5r/> .
@prefix skos:   <http://www.w3.org/2004/02/skos/core#> .

@prefix :     <http://data.europeana.eu/catalog#> .


:catalog a dcat:Catalog ;
  dct:description "The catalogue of all cultural heritage datasets aggregated by
Europeana"@en ;
  dct:issued "2009-12-10"^^xsd:date ;
  dct:license <http://creativecommons.org/publicdomain/zero/1.0/> ;
  dct:modified "2018-02-01"^^xsd:date ;
  dct:publisher <http://data.europeana.eu/organization/1482250000001617026> ;
  dct:spatial <http://publications.europa.eu/resource/authority/continent/EUROPE> ;
  dct:title "Europeana Datasets"@en ;
  dcat:themeTaxonomy <http://publications.europa.eu/resource/authority/data-theme> ;
  foaf:homepage <https://www.europeana.eu/> ;
  dcat:dataset [
      a dcat:Dataset ;
      dct:description "A dataset containing metadata descriptions of cultural
heritage digital objects from, and about, Europe. The dataset aggregates metadata
from more than 3,700 cultural heritage institutions."@en ;
      dct:publisher <http://data.europeana.eu/organization/1482250000001617026> ;
      dct:creator <http://data.europeana.eu/organization/1482250000001617026> ;
      dct:title "Europeana Aggregated Dataset of Cultural Heritage Digital
Objects"@en ;
      dcat:theme <http://publications.europa.eu/resource/authority/data-theme/EDUC>,
<http://publicdata.eu/def/category/education-culture-and-sport>;
      dct:conformsTo <http://www.europeana.eu/schemas/edm/> ;
      dcat:distribution [
      a dcat:Distribution ;
      dct:description "Data access via the Europeana SPARQL API"@en ;
      dcatap:availability <http://data.europa.eu/r5r/stable>;
      dct:conformsTo                          <http://www.europeana.eu/schemas/edm/>,
<http://www.w3.org/TR/sparql11-query/> ;
      adms:status <http://purl.org/adms/status/UnderDevelopment> ;
      dct:license <http://creativecommons.org/publicdomain/zero/1.0/> ;
      dcat:accessURL <https://pro.europeana.eu/page/sparql> ;
      dcat:accessService [
      a dcat:DataService;
```

```
      dct:description "The SPARQL API of Europeana APIs allows exploring connections
between Europeana data and outside data sources, like VIAF, Iconclass, Getty
Vocabularies (AAT), Geonames, Wikidata, and DBPedia. It allows one to delve into the
structured metadata of Europeana (for instance, to ask the question \"What are all
the French 18th-century painters with at least five artworks available through
Europeana\")"@en ;
      dct:title "Europeana SPARQL API" ;
      dcat:endpointURL <http://sparql.europeana.eu/>
      ] ;
      ]
  ] .

<http://creativecommons.org/publicdomain/zero/1.0/>
  a dct:LicenseDocument ;
  dct:type <http://purl.org/adms/licencetype/PublicDomain> .

<http://www.europeana.eu/schemas/edm/> a dct:Standard .
<https://www.europeana.eu/> a foaf:Document .
<http://purl.org/adms/status/UnderDevelopment> a dct:Standard .
<http://purl.org/adms/status/Completed> a dct:Standard .
<http://www.w3.org/TR/sparql11-query/> a dct:Standard .

<http://publications.europa.eu/resource/authority/continent/EUROPE> a dct:Location .
<http://publicdata.eu/def/category/education-culture-and-sport> a skos:Concept .


<http://data.europeana.eu/organization/1482250000001617026> a foaf:Agent ;
  dct:type <http://purl.org/adms/publishertype/NonProfitOrganisation> ;
  foaf:homepage <https://www.europeana.eu/> ;
  foaf:name "Europeana Foundation"@en, "Stichting Europeana"@nl .
```