

OPAL
OPEN DATA PORTAL

Deliverable D4.1

Vokabularspezifikation

Autoren: Afshin Amini, Zafar Habeeb Syed

Veröffentlichung	Öffentlich
Fälligkeitsdatum	30.09.2018
Fertigstellung	18.12.2018
Arbeitspaket	AP4
Typ	Bericht
Status	Final
Version	1.0

Kurzfassung:

Zur Darstellung von Metadaten aus unterschiedlichen Quellen im RDF-Format wird ein spezifisches Vokabular benötigt, um die einzelnen Eigenschaften von Datensätzen, Organisationen, Verteilung, Datenqualität u.v.a. semantisch abzubilden. In diesem Deliverable wird untersucht, welche Vokabulare es hierfür bereits gibt, die wiederverwendet und erweitert werden können. Anschließend wird die OPAL-Vokabularspezifikation vorgenommen. Diese wird um die Darstellung eines Beispieldatensatzes aus mCLOUD erweitert.

Schlagworte:

Vokabular, Darstellung, RDF, DCAT

Inhalt

Introduction	2
Goals and Non-Goals	2
State of the Art	2
Metadata Vocabulary Specification	4
3.1 Catalog	4
3.2 Dataset	5
3.3 Distribution	6
3.4 Quality metrics	7
Metadata Example for Sample Datasets	7
Conclusions	12

1 Introduction

One of the primary goal of OPAL is to extract metadata of datasets from different catalogs and enable easy access to Open Data. However, the distribution of datasets and their metadata are available in multiple format across catalogs. There is a need to convert all the collected metadata and represent them using a common specification, thereby enabling interoperability and discoverability of datasets across different catalogs. This document aims at defining a suitable vocabulary specification, i.e. an ontology to describe the metadata and other properties of the dataset. For the specified metadata, an RDF representation with existing and, if necessary, new vocabularies will be defined. Suitable existing vocabularies and their classes or attributes will be created using applications such as Linked Open Vocabularies (LOV) and analyzed based on their usability. At first, individual metadata for specific datasets and their RDF representation are recorded as training data. The training data serve as input data for a crawling/conversion component, which determines the assignment from the previously extracted metadata and its RDF representation. The document is organized as follows: Section 2 gives an overview of the existing vocabularies suitable to describe the metadata information and discuss their limitations. Section 3 details the list of metadata properties and the RDF vocabularies used for their representation. Finally, Section 4 provides examples of vocabulary specification for sample datasets.

1.1 Goals and Non-Goals

The primary goals of this document include:

1. Specification of a vocabulary for representing metadata of datasets handled by OPAL and metadata required by OPAL itself for managing the datasets
2. The specified vocabulary must reuse to the possible extent existing vocabularies. For example Dublin Core, void vocabulary etc.
3. Record for specific datasets the metadata records and their RDF properties to generate training examples for crawling component.
4. Finally, the vocabulary must conform to DCAT specification for converting the metadata to CKAN format (<https://github.com/ckan/ckanext-dcat>)

2 State of the Art

Several metadata vocabularies exist which can be used to represent information about datasets and effectively operate across various data catalogs published on the Web. Dublin Core¹ defines a vocabulary of generic properties which can be used to describe resources. These properties are designed to unify attributes of resources across different domains. For example, the information about the writer of a book, and the creator of a video can be specified by **dct:creator**, a Dublin Core property. However, due to their generic nature, standalone Dublin Core properties cannot be used to describe resources which demand specific descriptors. Dublin Core mitigates the lack of specificity by allowing other vocabularies to reuse and extend properties. For instance, DCAT² is an RDF vocabulary that incorporates properties from pre-existing vocabularies such as Dublin Core, and also additionally introduces new properties to describe and facilitate interoperability

¹ <http://dublincore.org/documents/dcmi-terms/>

² <https://www.w3.org/TR/vocab-dcat/>

and discoverability of datasets across data catalogs. DCAT specification consists of three main classes -- **Catalog**, **Dataset**, and **Distribution**. A Catalog is a collection of Datasets. A Dataset is a conceptual entity that represents the information published. A Distribution is the physical realization of a dataset and is made available for use, for example in the form of a file download or access via an API. A Dataset can have one or more Distributions. A data catalog conforms to DCAT if an RDF description of all the metadata of datasets and their distributions expressed using DCAT properties is available. Additionally, catalogs conforming to DCAT may include non-DCAT properties. Although, DCAT overcomes the problem of specificity by introducing additional properties for describing the datasets, however, it doesn't restrict the usage of properties. For instance, properties describing license and usage restrictions of a dataset may or may not be provided by the publisher. As a result, the potential users of the datasets may simply do not use it due to lack of license information. DCAT-AP³ is a customized specification of DCAT that adds constraints on usage of properties. DCAT-AP broadly classifies all the properties into three categories--mandatory, recommended and optional. For example, DCAT-AP recommends providing license information for the datasets and possibly at the distribution level. A catalog conforming to DCAT-AP also conforms to DCAT. For instance, govdata.de is a DCAT-AP compliant catalog. In govdata.de, all the metadata information about the datasets and their distributions is made available in the form of RDF description using DCAT. DCAT-AP has become the core model for describing and exchanging information across data portals. The DCAT-AP specifications provide the owners and publishers of datasets with a standardised description to store and publish data. In addition, it also helps organizations that aim to aggregate data from various catalogs and provide a single point of access to open data, thereby enabling the end consumers to easily find the datasets.

In [ADEQUATE](#), the authors describe the need for improved quality of the metadata and present a framework to monitor the quality of open data catalogs. [Zaveri et al.](#) presents a systematic review of approaches for measuring the quality of linked open data. The authors unify the quality aspects and provide a comprehensive list of quality dimensions and metrics. DQV⁴ is a vocabulary specification that extends the DCAT to describe the quality aspects of the datasets in a catalog. These quality specifications enable the end consumers of the datasets to better judge if the data presented is fit for use and establishes trust among the publishers, catalog providers and end users. DQV doesn't define any existing quality aspects, but enables the users of the vocabulary to define the quality measures. Each quality measure is an instance of **dqv:QualityMeasurement**, a metric value providing the qualitative or quantitative information about the datasets or distributions and defined by quality dimensions and quality metrics. A quality dimension is a characteristic of the dataset that is relevant to end consumers. A quality metric provides a concrete value to measure quality dimensions of datasets.

³ <https://www.dcat-ap.de/>

⁴ <https://www.w3.org/TR/vocab-dqv/>

3 Metadata Vocabulary Specification

In OPAL, we reuse existing vocabulary specifications and standardizations to store metadata information gathered from different catalogs. All the information collected is stored as RDF descriptions of Catalogs, Datasets, and Distributions using DCAT properties. Additionally, we create new resources for defining license information. The following subsections give detailed information about the metadata extracted and the relevant vocabulary that is used to represent the metadata.

3.1 Catalog

A data catalog is a repository of one or more datasets and is viewed as an instance of Catalog. In OPAL, the metadata of datasets from several data catalogs is extracted. In addition to providing the RDF description of the datasets, the information about the catalog is extracted and stored in RDF format. Given a data catalog, the following information is extracted

- **Title:** The given name of the catalog
- **Description:** A short textual description of the catalog
- **Language:** The language used for the description of metadata of the datasets, for example, the language for describing title, description, and etc. Note that, a catalog can have multiple values for Language if different datasets have a different language
- **Homepage:** The link to the start page of the catalog
- **Publisher:** An individual/organization responsible for maintaining and publishing the catalog online
- **Geo-Location:** Represents the geographical area covered by the catalog
- **Issued:** The formal date on which the catalog was made online
- **Modified:** The most recent date on which the catalog was updated, modified or changed
- **License:** Link to formal documents specifying the conditions under which the catalog can be used or modified
- **Dataset:** A dataset which is part of the catalog
- **Format:** The format for distribution of metadata

The following table gives an overview of the RDF properties used to represent the metadata:

Metadata	RDF-Property	Range
Title	< http://purl.org/dc/terms/title >	xsd:String
Description	< http://purl.org/dc/terms/description >	xsd:String
Language	< http://purl.org/dc/terms/language >	< http://purl.org/dc/terms/LinguisticSystem >
Homepage	< http://xmlns.com/foaf/0.1/homepage >	< http://xmlns.com/foaf/0.1/Document >
Publisher	< http://purl.org/dc/terms/publisher >	< http://xmlns.com/foaf/0.1/Agent >

Geo-Location	< http://purl.org/dc/terms/spatial >	< http://purl.org/dc/terms/Location >
Issued	< http://purl.org/dc/terms/issued >	xsd:dateTime
Modified	< http://purl.org/dc/terms/modified >	xsd:dateTime
License	< http://purl.org/dc/terms/license >	URL
Dataset	< http://www.w3.org/ns/dcat#dataset >	< http://www.w3.org/ns/dcat#Dataset >
Format	< http://purl.org/dc/terms/format >	< http://purl.org/dc/terms/MediaTypeOrExtent >

Table 1: OPAL vocabulary for specifying catalogs

For all the properties, the **domain** is a unique URI generated for each catalog and having type <<http://www.w3.org/ns/dcat#Catalog>>. For each catalog, a unique identifier is assigned and the URI is generated by appending the identifier to the base URI <<http://mcloud.projekt-opal.de/catalog>>.

3.2 Dataset

Each dataset in a given data catalog is an instance of class Dataset. For each dataset, the following metadata is extracted:

- **Title:** The given title of the dataset
- **Description:** Free-text by an individual/organization used to describe the dataset
- **Date Issued:** Date on which the dataset was first published
- **Date Modified:** The most recent date on which the dataset was updated, changed or modified
- **Language:** The language used to describe the metadata of the dataset. Note that, the metadata can have a value from the list of values available for the catalog
- **Publisher:** An entity responsible for making the dataset available
- **Creator:** An individual/organization responsible for creating and maintaining the dataset
- **Accrual Period:** The time-frequency at which the dataset is updated
- **Geo-Location:** The geographical region covered by the dataset
- **Temporal Period:** The time frame that the dataset covers
- **Category:** Every data catalog has a set of predefined categories. This is the value of the given category to which the dataset belongs
- **Contact Point:** An individual/organization
- **Landing Page:** A unique URL to access the dataset information
- **Distribution:** Link to a distribution through which the dataset is made available.

Metadata	RDF-Property	Range
Title	< http://purl.org/dc/terms/title >	xsd:String
Description	< http://purl.org/dc/terms/description >	xsd:String

Date Issued	< http://purl.org/dc/terms/issued >	xsd:dateTime
Date Modified	< http://purl.org/dc/terms/modified >	xsd:dateTime
Language	< http://purl.org/dc/terms/language >	< http://purl.org/dc/terms/LinguisticSystem >
Publisher	< http://purl.org/dc/terms/publisher >	< xmlns.com/foaf/0.1/Agent >
Creator	< http://purl.org/dc/terms/creator >	< xmlns.com/foaf/0.1/Agent >
Accrual Period	< http://purl.org/dc/terms/accrualPeriodicity >	< purl.org/dc/terms/Frequency >
Geo-Location	< http://purl.org/dc/terms/spatial >	< purl.org/dc/terms/Location >
Temporal Period	< http://purl.org/dc/terms/temporal >	< purl.org/dc/terms/PeriodOfTime >
License	< http://purl.org/dc/terms/license >	URL
Category	< www.w3.org/ns/dcat#theme >	< www.w3.org/2004/02/skos/core#Concept >
Contact Point	< www.w3.org/ns/dcat#contactPoint >	< www.w3.org/2006/vcard/ns#Organization >
Landing Page	< www.w3.org/ns/dcat#landingPage >	< xmlns.com/foaf/0.1/Document >
Distribution	< www.w3.org/ns/dcat#distribution >	< www.w3.org/ns/dcat#Distribution >

Table 2: OPAL vocabulary for specifying datasets

Similar to catalogs, unique URI's are generated to represent the found datasets and used as domain values for the properties described. The base URI for the the datasets used is <<http://mcloud.projekt-opal.de/dataset>>.

3.3 Distribution

A distribution is a specific representation in which the dataset is made available. For example, in the form of a file download or an end-point. A dataset may have one or more distributions each of which is an instance of Distribution class. The following metadata is extracted for each distribution:

- **Title:** The given title of the distribution
- **Description:** Short text describing the distribution
- **Date Issued:** The first date of issuance of the distribution
- **Date Modified:** The last recent date on which the distribution is modified
- **License:** The link to document dictating the terms and conditions on which the distribution can be used or modified
- **Access URL:** The link to end-point, feed where the distribution can be accessed

- **Download URL:** The link to file download in case if available.

The following Table gives the list of RDF properties used to represent the metadata. The domain values for these properties are unique URI's generated for each distribution of a given dataset. The URI's are generated by appending the unique identifier of the distribution to the base URL <<http://mcloud.projekt-opal.de/distribution>>.

Metadata	RDF-Property	Range
Title	< http://purl.org/dc/terms/title >	xsd:String
Description	< http://purl.org/dc/terms/description >	xsd:String
Date Issued	< http://purl.org/dc/terms/issued >	xsd:dateTime
Date Modified	< http://purl.org/dc/terms/modified >	xsd:dateTime
License	< http://purl.org/dc/terms/license >	URL
Access URL	< http://www.w3.org/ns/dcat#accessURL >	< http://www.w3.org/2000/01/rdf-schema#Resource >
Download URL	< http://www.w3.org/ns/dcat#downloadURL >	< http://www.w3.org/2000/01/rdf-schema#Resource >

Table 3: OPAL vocabulary for specifying distributions

3.4 Quality metrics

In addition to extracting the metadata of the datasets from various catalogs, a wide range of metrics are defined for determining the quality of the datasets. These metrics are defined to evaluate the quality aspects of the datasets across various dimensions. Broadly, these dimensions are categorized into four categories and are defined as the type of RDF class **dqv:Category**. Each category consists of one or more quality dimensions represented as **dqv:Dimension** and are linked to the category using the RDF-Property **dqv:inCategory**. A quality dimension is used to represent criteria relevant for assessing quality of the datasets. These representations are realized by one or more quality metrics that assigns concrete values to these dimensions. A more detailed information about the definitions and usage of these quality measures can be found in the upcoming deliverable D3.2.

4 Metadata Example for Sample Datasets

mcloud.de is a catalog that hosts several datasets in 6 pre-defined categories. for example by using the opal vocabulary and for 3 datasets from D1.2 (49, 84, 378) the RDF is like this:

NOTE: the offline HTML pages and also RDF generated for each page is available at GitHub page: <https://github.com/projekt-opal/vocabularyExample>

RDF-SUBJECT	RDF-PROPERTY	RDF-OBJECT
< http://projekt-opal.de/dataset/bundfar >	< http://www.w3.org/1999/02/2 >	< http://www.w3.org/ns/dcat# >

D4.1 - Vokabularspezifikation

breiefdeswasserlaufs>	2-rdf-syntax-ns#type>	Dataset>
<http://projekt-opal.de/dataset/bundfarbreiefdeswasserlaufs>	<http://purl.org/dc/terms/title>	"Bund: Farbreief des Wasserlaufs"
<http://projekt-opal.de/dataset/bundfarbreiefdeswasserlaufs>	<http://purl.org/dc/terms/description>	"Farbreief des Wasserlaufs"
<http://projekt-opal.de/dataset/bundfarbreiefdeswasserlaufs>	<http://purl.org/dc/terms/issued>	"14.09.2016"
<http://projekt-opal.de/dataset/bundfarbreiefdeswasserlaufs>	<http://www.w3.org/ns/dcat#theme>	<http://projekct-opal.de/theme/mcloud/waterwaysAndWaters>
<http://projekt-opal.de/dataset/bundfarbreiefdeswasserlaufs>	<http://www.w3.org/ns/dcat#distribution>	<http://projekt-opal.de/distribution/bundfarbreiefdeswasserlaufs_0>
<http://projekt-opal.de/dataset/bundfarbreiefdeswasserlaufs>	<http://www.w3.org/ns/dcat#distribution>	<http://projekt-opal.de/distribution/bundfarbreiefdeswasserlaufs_1>
<http://projekt-opal.de/dataset/bundfarbreiefdeswasserlaufs>	<http://www.w3.org/ns/dcat#distribution>	<http://projekt-opal.de/distribution/bundfarbreiefdeswasserlaufs_2>
<http://projekt-opal.de/dataset/bundfarbreiefdeswasserlaufs>	<http://purl.org/dc/terms/publisher>	<http://projekt-opal.de/agent/bundfarbreiefdeswasserlaufs>
<http://projekt-opal.de/dataset/bundfarbreiefdeswasserlaufs>	<http://www.w3.org/ns/dcat#landingPage>	<https://mcloud.de/web/guest/suche/-/results/detail/bundfarbreiefdeswasserlaufs?_mcloudsearchportlet_backURL=https%3A%2F%2Fmcloud.de%2Fweb%2Fguest%2Fsuche%2F-%2Fresults%2FsearchAction%3F_mcloudsearchportlet_page%3D74%26_mcloudsearchportlet_sort%3Dlatest>
<http://projekct-opal.de/theme/mcloud/waterwaysAndWaters>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	"skos:Concept"
<http://projekct-opal.de/theme/mcloud/waterwaysAndWaters>	<https://www.w3.org/2000/01/rdf-schema#label>	"Wasserstraßen und Gewässer"
<http://projekt-opal.de/distribution/bundfarbreiefdeswasserlaufs_0>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/ns/dcat#Distribution>
<http://projekt-opal.de/distribution/bundfarbreiefdeswasserlaufs_0>	<http://purl.org/dc/terms/license>	<http://www.gesetze-im-internet.de/geonutzv/index.html>
<http://projekt-opal.de/distribution/bundfarbreiefdeswasserlaufs_0>	<http://www.w3.org/ns/dcat#downloadURL>	<http://atlas.wsv.bund.de/inspire/atom/hoehen.xml>
<http://projekt-opal.de/distribution/bundfarbreiefdeswasserlaufs_1>	<http://www.w3.org/ns/dcat#accessURL>	<http://atlas.wsv.bund.de/inspire/el/wms?REQUEST=GetCapabilities&SERVICE=wms>

D4.1 - Vokabularspezifikation

<http://projekt-opal.de/distribution/bundfarbreliefdeswasserlaufs_1>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/ns/dcat#Distribution>
<http://projekt-opal.de/distribution/bundfarbreliefdeswasserlaufs_1>	<http://purl.org/dc/terms/licence>	<http://www.gesetze-im-internet.de/geonutzv/index.html>
<http://projekt-opal.de/distribution/bundfarbreliefdeswasserlaufs_2>	<http://www.w3.org/ns/dcat#accessURL>	<http://atlas.wsv.bund.de/inspire/el/wms?service=wms&request=getcapabilities>
<http://projekt-opal.de/distribution/bundfarbreliefdeswasserlaufs_2>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/ns/dcat#Distribution>
<http://projekt-opal.de/distribution/bundfarbreliefdeswasserlaufs_2>	<http://purl.org/dc/terms/licence>	<http://www.gesetze-im-internet.de/geonutzv/index.html>
<http://projekt-opal.de/agent/bundfarbreliefdeswasserlaufs>	<http://xmlns.com/foaf/0.1/name>	"Informationstechnikzentrum Bund (ITZBund)"
<http://projekt-opal.de/agent/bundfarbreliefdeswasserlaufs>	<http://xmlns.com/foaf/0.1/homepage>	<https://www.itzbund.de>
<http://projekt-opal.de/agent/bundfarbreliefdeswasserlaufs>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://xmlns.com/foaf/0.1/Agent>
<http://projekt-opal.de/dataset/digitalebundeswasserstraenkarteimmastab11000000>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/ns/dcat#Dataset>
<http://projekt-opal.de/dataset/digitalebundeswasserstraenkarteimmastab11000000>	<http://purl.org/dc/terms/title>	"Digitale Bundeswasserstraßenkarte im Maßstab 1:1.000.000"
<http://projekt-opal.de/dataset/digitalebundeswasserstraenkarteimmastab11000000>	<http://purl.org/dc/terms/description>	"Digitale Bundeswasserstraßenkarte im Maßstab 1:1.000.000"
<http://projekt-opal.de/dataset/digitalebundeswasserstraenkarteimmastab11000000>	<http://www.w3.org/ns/dcat#landingPage>	<https://mcloud.de/web/guest/suche/-/results/detail/digitalebundeswasserstraenkarteimmastab11000000?_mcloudsearchportlet_backURL=https%3A%2F%2Fmcloud.de%2Fweb%2Fguest%2Fsuche%2F-%2Fresults%2FsearchAction%3F_mcloudsearchportlet_page%3D73%26_mcloudsearchportlet_sort%3Dlatest>
/digitalebundeswasserstraenkarteimmastab11000000>		
<http://projekt-opal.de/dataset/digitalebundeswasserstraenkarteimmastab11000000>	<http://www.w3.org/ns/dcat#theme>	<http://projekt-opal.de/theme/mcloud/waterwaysAndWaters>
<http://projekt-opal.de/dataset/digitalebundeswasserstraenkarteimmastab11000000>	<http://purl.org/dc/terms/issued>	"14.09.2016"
<http://projekt-opal.de/dataset/digitalebundeswasserstraenkarteimmastab11000000>	<http://www.w3.org/ns/dcat#di	<http://projekt-opal.de/distribu

D4.1 - Vokabularspezifikation

undeswasserstraenkarteimmastab11000000>	tribution>	tion/digitalebundeswasserstraenkarteimmastab11000000_0>
<http://projekt-opal.de/dataset/digitalebundeswasserstraenkarteimmastab11000000>	<http://purl.org/dc/terms/publisher>	<http://projekt-opal.de/agent
<http://projექt-opal.de/theme/mcloud/waterwaysAndWaters>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	"skos:Concept"
<http://projექt-opal.de/theme/mcloud/waterwaysAndWaters>	<https://www.w3.org/2000/01/rdf-schema#label>	"Wasserstraßen und Gewässer"
<http://projekt-opal.de/distribution/digitalebundeswasserstraenkarteimmastab11000000_0>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/ns/dcat#Distribution>
<http://projekt-opal.de/distribution/digitalebundeswasserstraenkarteimmastab11000000_0>	<http://www.w3.org/ns/dcat#accessURL>	<http://atlas.wsv.bund.de/dbwk1000/wms?REQUEST=GetCapabilities&SERVICE=WMS>
<http://projekt-opal.de/distribution/digitalebundeswasserstraenkarteimmastab11000000_0>	<http://purl.org/dc/terms/license>	<http://www.gesetze-im-internet.de/geonutzv/index.html>
<http://projekt-opal.de/agent/digitalebundeswasserstraenkarteimmastab11000000>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://xmlns.com/foaf/0.1/Agent>
<http://projekt-opal.de/agent/digitalebundeswasserstraenkarteimmastab11000000>	<http://xmlns.com/foaf/0.1/name>	"Generaldirektion Wasserstraßen und Schifffahrt (GDWS)"
<http://projekt-opal.de/agent/digitalebundeswasserstraenkarteimmastab11000000>	<http://xmlns.com/foaf/0.1/homepage>	<http://www.wsv.de/>
<http://projekt-opal.de/dataset/radwegegishamburg>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/ns/dcat#Dataset>
<http://projekt-opal.de/dataset/radwegegishamburg>	<http://purl.org/dc/terms/title>	"RadwegeGis Hamburg"
<http://projekt-opal.de/dataset/radwegegishamburg>	<http://purl.org/dc/terms/description>	"Dieser Datensatz ist und war nie vollständig und wird nicht mehr gepflegt. Die Inhalte sind in verschiedene Einzeldatensätze zum Thema Radverkehr aufgeteilt."
<http://projekt-opal.de/dataset/radwegegishamburg>	<http://purl.org/dc/terms/issued>	"07.09.2014"
<http://projekt-opal.de/dataset/radwegegishamburg>	<http://www.w3.org/ns/dcat#theme>	<http://projექt-opal.de/theme/mcloud/street>
<http://projekt-opal.de/dataset/radwegegishamburg>	<http://www.w3.org/ns/dcat#distribution>	<http://projekt-opal.de/distribution/radwegegishamburg_0>

D4.1 - Vokabularspezifikation

<http://projekt-opal.de/dataset/radwegegishamburg>	<http://www.w3.org/ns/dcat#distribution>	<http://projekt-opal.de/distribution/radwegegishamburg_1>
<http://projekt-opal.de/dataset/radwegegishamburg>	<http://purl.org/dc/terms/publisher>	<http://projekt-opal.de/agent/radwegegishamburg>
<http://projekt-opal.de/dataset/radwegegishamburg>	<http://www.w3.org/ns/dcat#landingPage>	<https://mcloud.de/web/guest/suche/-/results/detail/radwegegishamburg?_mcloudsearchportlet_backURL=https%3A%2F%2Fmcloud.de%2Fweb%2Fguest%2Fsuche%2F-%2Fresults%2FsearchAction%3F_mcloudsearchportlet_page%3D1%26_mcloudsearchportlet_sort%3Dlatest>
<http://projekt-opal.de/theme/mcloud/street>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	"skos:Concept"
<http://projekt-opal.de/theme/mcloud/street>	<https://www.w3.org/2000/01/rdf-schema#label>	"Straßen"
<http://projekt-opal.de/distribution/radwegegishamburg_0>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/ns/dcat#Distribution>
<http://projekt-opal.de/distribution/radwegegishamburg_0>	<http://purl.org/dc/terms/license>	<https://www.govdata.de/dl-die/by-2-0>
<http://projekt-opal.de/distribution/radwegegishamburg_0>	<http://www.w3.org/ns/dcat#downloadURL>	<http://daten-hamburg.de/transport_verkehr/radwegeGIS/RadwegeGIS_HH_2014-09-17.zip>
<http://projekt-opal.de/distribution/radwegegishamburg_1>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/ns/dcat#Distribution>
<http://projekt-opal.de/distribution/radwegegishamburg_1>	<http://purl.org/dc/terms/license>	<https://www.govdata.de/dl-die/by-2-0>
<http://projekt-opal.de/distribution/radwegegishamburg_1>	<http://www.w3.org/ns/dcat#downloadURL>	<http://geoportal-hamburg.de/Geoportal/geo-online/>
<http://projekt-opal.de/agent/radwegegishamburg>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://xmlns.com/foaf/0.1/Agent>
<http://projekt-opal.de/agent/radwegegishamburg>	<http://xmlns.com/foaf/0.1/name>	"Hamburg: Behörde für Wirtschaft, Verkehr und Innovation, Amt für Verkehr und Straßenwesen"
<http://projekt-opal.de/agent/radwegegishamburg>	<http://xmlns.com/foaf/0.1/homepage>	<http://www.hamburg.de/bwvi/verkehr-strassenwesen/>

Table 4: Example use of the OPAL vocabulary for representing mCLOUD datasets

5 Conclusions

The vocabulary used in OPAL extends existing suitable vocabularies, most notably DCAT, for representing the metadata extracted from different sources. Entities that can be represented include catalogs, datasets, distributions, and quality metrics. After its specification, it has been tested to work with existing transformation processes required for providing a CKAN API (see architecture in deliverable D1.3). This makes sure that the representation conforms to agreed existing vocabulary.

For generating RDF from crawled metadata, there is still the issue of minting URIs for the respective resources. The example metadata shown in section 4 illustrates one such approach. During the scope of the project, this method might change depending on further requirements due to the interlinking and fusion processes (see Figure 1 for an example), as well as making the metadata dereferenceable.

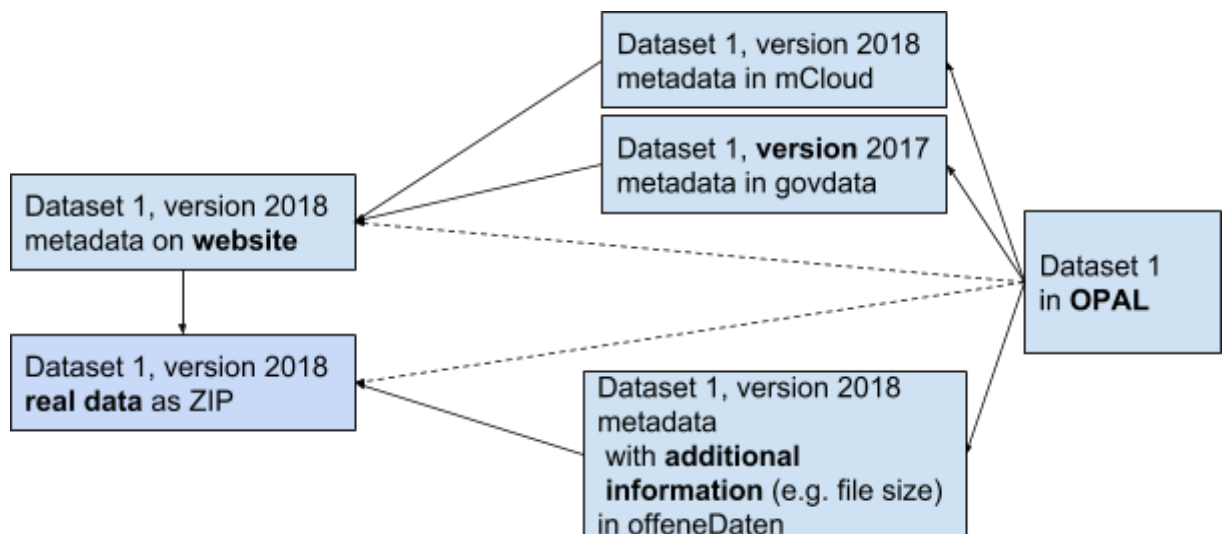


Figure 1: Example fusion process based on extracted dataset metadata from different sources