



**KDI** ● **Knowledge and Data Integration**

**DCAT**

Metadata & Catalogs

**W6.L11.M4.T11.1.2**

# Contents

**1** Top level view

**2** DCAT

# Contents

**1** Top level view

2 DCAT

# Top level view

When we are looking for something on the internet we usually rely on search engines. When we have data?

- Normal techniques may not work (i.e. encoded files);
- what we look for is not in the data itself (i.e. time pertinence of the data);
- often data are not linked and it needs additional insight to understand what the data are representing.

Because of this we need special tools to handle this problems.

# Catalogs

*Catalogs* are tools used to store and retrieve data in a sustainable way.

- Perform queries on the *data & metadata*;
- most of them allow to store data in a decentralised way (GDPR);
- allow to have access control to resources;
- allow to implement more complex metrics on usage of data.

## Example of instances

- <https://dati.trentino.it/>
- <https://data.gov.uk>

## Example of platforms

- CKAN
- DKAN
- Socrata

# Metadata

When we share data we need to share also it's *metadata* together. Some format can contains some metadata themselves but often we *infer* them.

## Example

John Doe	129343	30
Jane Doe	131231	25

**Table:** student-score-2019.csv

# Metadata

When we share data we need to share also it's *metadata* together. Some format can contains some metadata themselves but often we *infer* them.

## Example

Even if we do have column's headers

Name	Mat.	Score
John Doe	129343	30
Jane Doe	131231	25

**Table:** student-score-2019.csv

# Metadata

We can define this as *plain text*, but that is useful for humans. Moreover natural language is ambiguous by its nature, and difficult to be read by a machine.

There is a need for a more *formal language* to express this kind of metadata.



# Data Catalog Vocabulary (DCAT)

It is a **W3C standard** vocabulary used to define formally data catalogs together with metadata. Currently in version 2.0; you can reference to this URL for details:

<https://www.w3.org/TR/vocab-dcat-2/>

## Note

Some parts will not be discussed in depth because concepts related to them will be expanded in future lessons.

# DCAT 2

Principal element in DCAT2 are:

Catalog

collection of metadata about *Dataset* & *DataService*

Dataset

collection of data, published or curated by a single agent;

Distribution

represents an accessible form of a *Dataset* such as a downloadable file;

DataService

collection of operations accessible through an interface (API) that provide access to one or more *Datasets*;

CatalogRecord

represents a metadata item in the *Catalog*, primarily concerning the registration information, i.e. *uploader*.

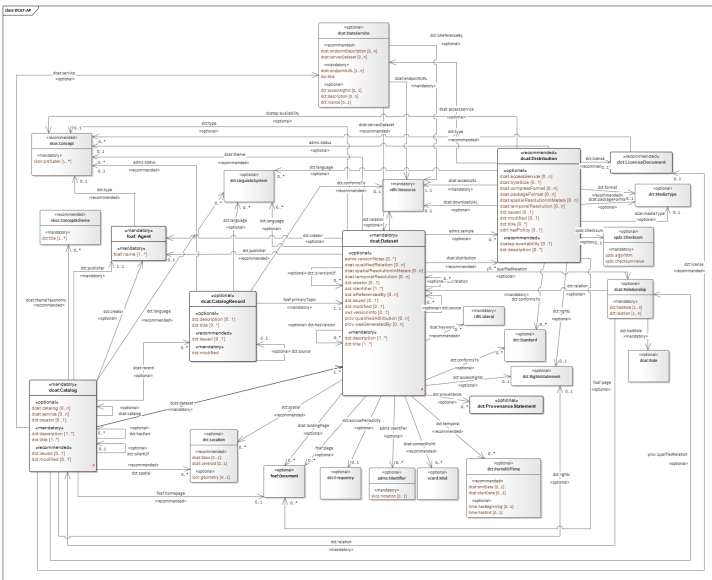
# DCAT-Application Profile

The European Commission's Interoperability Solutions for European Public Administrations, in the attempt at improving semantic interoperability in European eGovernment systems, pushed a new portal:

<https://data.europa.eu/euodp/en/home>

This portal aims to interconnect all data coming from different public (and private) institutions. To do so they created an *application profile*, based on DCAT, to allow data&metadata exchange.

# DCAT-AP 2.0.1



# DCAT-AP - Extensions

- National extensions
- GeoDCAT-AP
- StatDCAT-AP

## Why is important?

- Visibility
- Allows complex queries to catalogs
- Enable automation



KDI : Knowledge and Data Integration



**W6.L11.M4.T11.1.2**



**DCAT**

Metadata & Catalogs