

URBAN CO-CREATION DATA LAB

Open Data and Metadata

António Soares - AMA

10 February 2022



Co-financed by the Connecting Europe Facility of the European Union

Index

1. Introduction to AMA
2. Open Data
3. dados.gov
4. Data quality
5. Metadata
6. DCAT; DCAT-AP



Introduction to AMA

AMA

ADMINISTRATIVE MODERNIZATION AGENCY

Public institute created in 2007. AMA follows the Presidency of the Council of Ministers' responsibilities in the areas of **modernization, administrative simplification** and **electronic government**, under superintendence and tutelage of the Secretary of State Assistant and of Administrative Modernization. The Agency's activity is divided in three major axes:

Service Delivery, Digital Transformation and Simplification.

Open Data

Open Data Economic Value

The European Union identifies open data as a **pillar** for the development of the European economy and predicts that the **economic value derived from its reuse** will quadruple between 2018 and 2030, estimating that by then it will reach a value of **194 billion euros in 2030**.

Open Data

The Open Definition sets out principles that define “openness” in relation to **data and content.**

“Open means **anyone** can **freely access, use, modify, and share** for **any purpose**(subject, at most, to requirements that preserve provenance and openness).”

Open data represents a very important subset of the vast domain of public sector info

and is part of policies dedicated to Open Government, combining principles of

transparency,

democracy,

participation

and collaboration.

dados.gov

dados.gov – National Open Data Portal

It was developed (2011) and managed by AMA and is based on the open-source platform uData.



Central catalog of open data in Portugal

Its function is to:

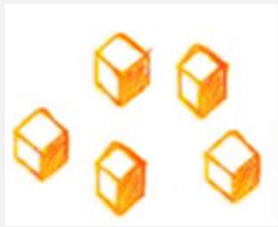
- aggregate, reference and enlist open data from different domains and public administration systems (private sector);
- serve as an indexing portal for other portals / open data catalogs.

Lei nº68/2021 -> Nat. Law that regulates the dados.gov role and transposes the EU directive.

dados.gov

"data that anyone can access, use, modify and share, for any purpose".
"came as you are"

Public Administration



Open Data Portal
dados.gov



Community of Reusers



Citizens



Disclosure and Availability

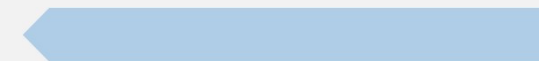


Sharing

Consumption



Value



Services and Platforms

Transparency, Information, Trust and Accountability

dados.gov

National Open Data Portal - numbers

ABRIR, PARTILHAR, CONSTRUIR

O catálogo central de dados abertos em Portugal

Bem-vindo ao dados . gov +, a versão melhorada do portal nacional de dados abertos. Caso já seja um reutilizador de dados, consulte a área de Documentação ou entre em Contacto connosco.

VER DADOS

121

organizations

2431

users

4835

datasets

51

reuses

120k visits last 24m **65k** v. l. 12m

Analytic
Models



Data quality

Best practices and recommendations for high quality publication

Data quality covers different aspects, for example **consistency, conformity, completeness or documentation**.

Help reusers to easily discover, reuse, analyse, visualise or process data.

The FAIR guiding principles for scientific data management provide a framework for grouping the different aspects of data quality.

Grouped by 4 dimensions:

- **Findability (Localizáveis);**
- **Accessibility (Acessíveis);**
- **Interoperability (Interoperáveis);**
- **Reusability (Reutilizáveis).**

Metadata



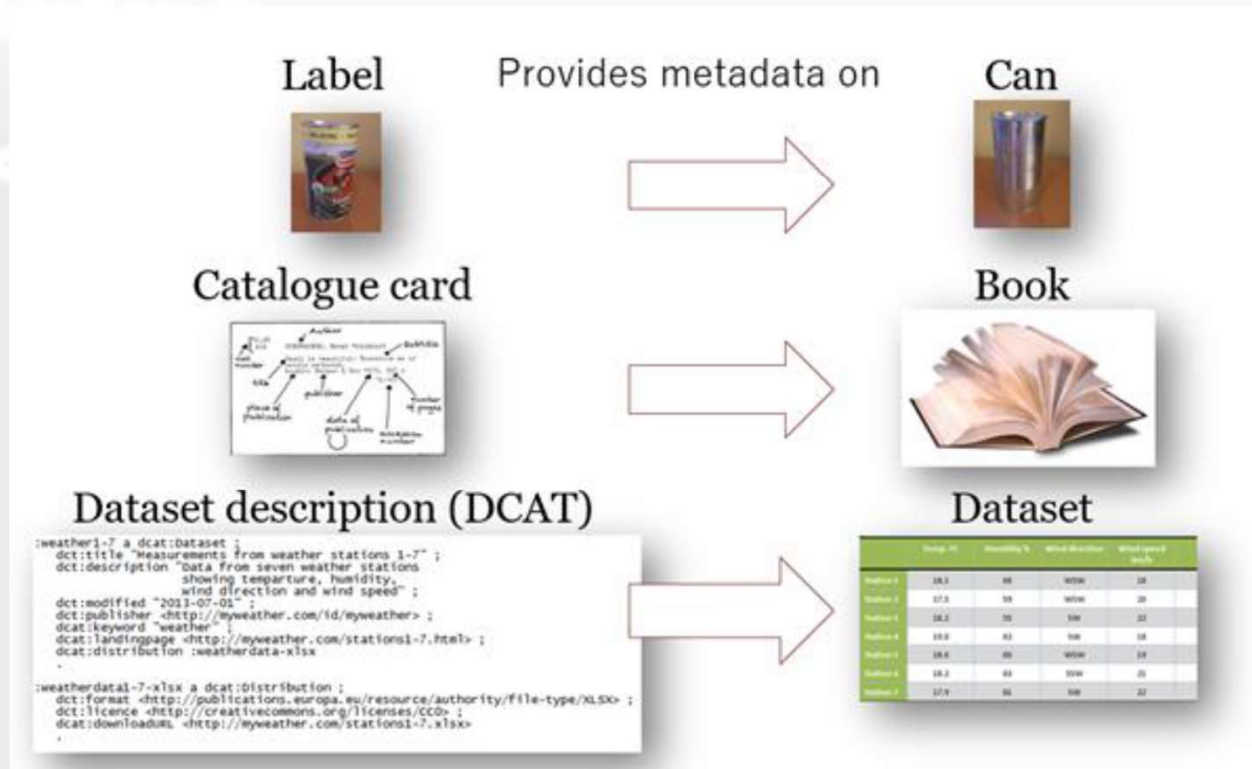
To be Findable: (meta)data are registered or indexed in a **searchable resource**.

To be accessible: (meta)data are retrievable by their identifier using a standardized communications protocol.

To be interoperable: (meta)data use a formal, accessible, shared, and broadly applicable language for **knowledge representation**.

To be re-usable: meta(data) have a plurality of accurate and **relevant attributes**.

Metadata



- Data that provide [information about other data](#).
- Metadata [summarizes basic information about data](#), making finding & working with particular instances of data easier.

DCAT-AP

The Data Catalog Vocabulary ([DCAT](#)) is a vocabulary designed to facilitate **interoperability** between data catalogs published on the Web.

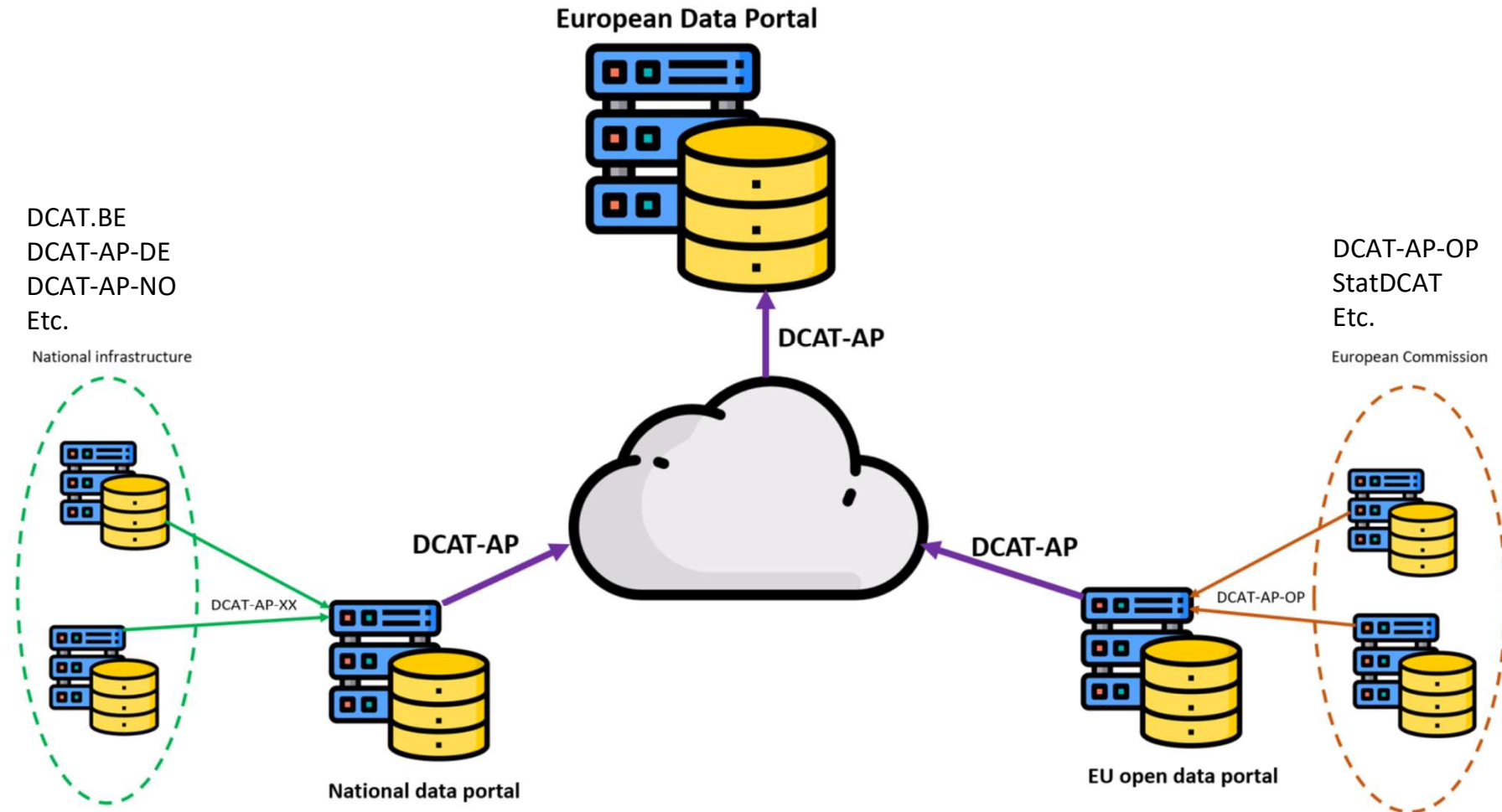
- DCAT enables a publisher to **describe** datasets (and data services) in a catalog using a **standard model and vocabulary** that facilitates the consumption and aggregation of metadata from multiple catalogs.

The DCAT Application Profile ([DCAT-AP](#)) for data portals in Europe is a specification for **describing public sector datasets in Europe**.

- Its basic use case **is to enable a cross-data portal search** for data sets and make public sector data better searchable **across borders and sectors**.

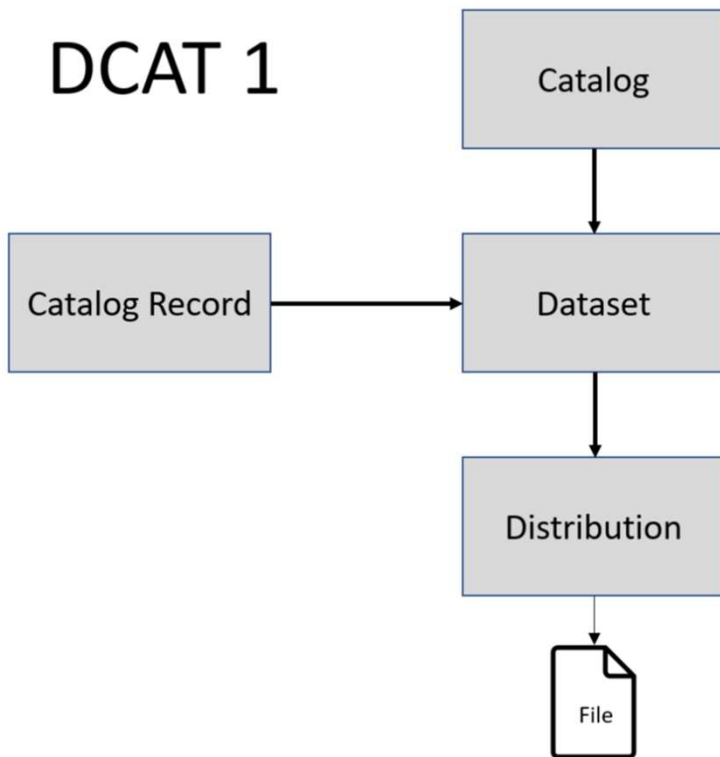
DCAT-AP

Data Catalogue interoperability



DCAT-AP

Main entities in DCAT



Catalog: curated collection of metadata about datasets.

Dataset: collection of data, published or curated by a single agent, and available for access or download in one or more formats.

Catalog Record: record in a data catalog, describing a single dataset

Distribution: a specific available form of a dataset.

All these contain metadata stored in a database

The **actual data is a file in a filestore.**

DCAT-AP

Dataset

<i>Title</i>	COVID-19 Coronavirus data (en), Données relatives au coronavirus COVID-19 (fr)
<i>Contact point</i>	E-mail ECDC
<i>Theme</i>	Health
<i>Publisher</i>	ECDC
<i>Language</i>	English
<i>Update frequency</i>	Weekly
<i>Distribution</i>	JSON, XML

```
e:dataset-001 a dcat:Dataset ;  
  
dct:title "COVID-19 Coronavirus data"@en ;  
dct:title "Données relatives au coronavirus COVID-19"@fr ;  
dcat:contactPoint [ a v: Kind ; v:hasEmail <mailto:info@ecdc.europa.eu> ; ] ;  
dcat:distribution e:dist-001 ;  
dcat:distribution e:dist-002 ;  
dct:publisher  
  <http://publications.europa.eu/resource/authority/corporate-body/ECDC> ;  
dcat:theme <http://publications.europa.eu/resource/authority/data-theme/HEAL> ;  
dct:accrualPeriodicity  
  <http://publications.europa.eu/resource/authority/frequency/WEEKLY> ;  
.
```

Resource Description Framework RDF
Graph representation of Knowledge

ld-json

rdf+xml

turtle

...

DCAT-AP

Dataset

<i>Title</i>	COVID-19 Coronavirus data (en), Données relatives au coronavirus COVID-19 (fr)
<i>Contact point</i>	E-mail ECDC
<i>Theme</i>	Health
<i>Publisher</i>	ECDC
<i>Language</i>	English
<i>Update frequency</i>	Weekly
<i>Distribution</i>	JSON, XML

```
e:dataset-001 a dcat:Dataset ;

dct:title "COVID-19 Coronavirus data"@en ;
dct:title "Données relatives au coronavirus COVID-19"@fr ;
dcat:contactPoint [ a v: Kind ; v:hasEmail <mailto:info@ecdc.europa.eu> ; ] ;
dcat:distribution e:dist-001;
dcat:distribution e:dist-002 ;
dct:publisher
  <http://publications.europa.eu/resource/authority/corporate-body/ECDC> ;
dcat:theme <http://publications.europa.eu/resource/authority/data-theme/HEAL> ;
dct:accrualPeriodicity
  <http://publications.europa.eu/resource/authority/frequency/WEEKLY> ;
.
```

Distribution

<i>Description</i>	Data on the geographic distribution of COVID-19 cases worldwide
<i>Licence</i>	CC-BY-4.0
<i>Access URL</i>	https://opendata.ecdc.europa.eu/covid19/nationalcasedeath/json
<i>Format</i>	JSON

```
a dcat:Distribution ;
dcat:accessURL
  <https://opendata.ecdc.europa.eu/covid19/nationalcasedeath/json> ;
dct:description "Data on the geographic distribution of COVID-19 cases worldwide"@en ;
dct:format
  <http://publications.europa.eu/resource/authority/file-type/JSON> ;
dct:license
  <http://publications.europa.eu/resource/authority/licence/CC_BY_4_0> ;
.
```

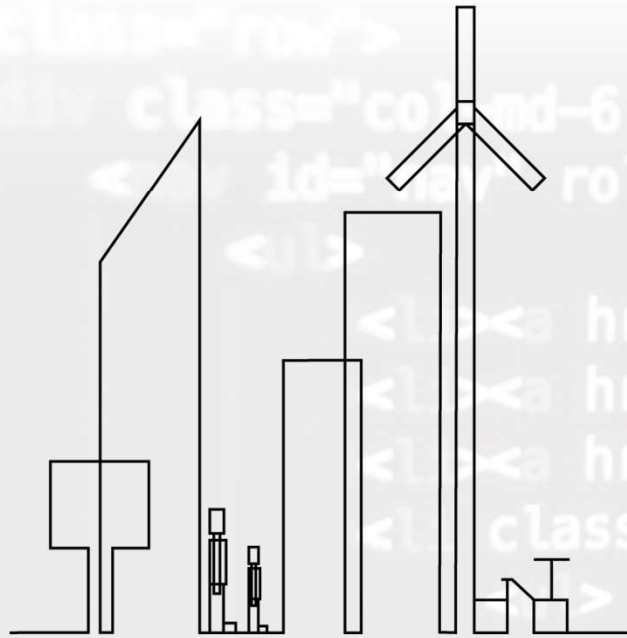

Thank you

antonio.soares@ama.pt

Direção de Estratégias e Governo Digital



Co-financed by the Connecting Europe
Facility of the European Union



URBAN CO-CREATION DATA LAB

Metadata JSON-LD Generator Tool



Co-financed by the Connecting Europe Facility of the European Union

Milestone 8 >

Publication of the relevant datasets, resulting from the Action on a national portal or catalogue harvested by the European Data Portal

- Compliance with the Metadata Quality Assessment (MQA) tool for datasets
 - Relevant datasets are published on a national portal or catalogue that is harvested by the European Data Portal
 - MQA can be performed

Problem

CML don't have their metadata in Lisbon Open Data Portal in a format readable by the MQA tool

- application/ld+json
- application/rdf+xml
- application/n-triples
- application/trig
- text/turtle
- text/n3
- text/plain

Metadata files

How can the metadata of the open datasets used in the project comply with the MQA tool?

- Complies with DCAT-AP and DCAT-AP derivatives
- Disclosure of information to which DCAT-AP is not obligated
- Accessibility of the data referenced in the metadata through the Access and Download URL
- Machine readability of the referenced data
- Use of licenses

Metadata creation application

Was created an app that generates json-ld metadata files that are fully compliant with the MQA

<https://metadata-app-eu.herokuapp.com>

Thank you

<http://urbandatalab.pt/>



Co-financed by the Connecting Europe
Facility of the European Union