



European
Commission



Joint Research Centre

the European Commission's
in-house science service

Using DCAT-AP for research data

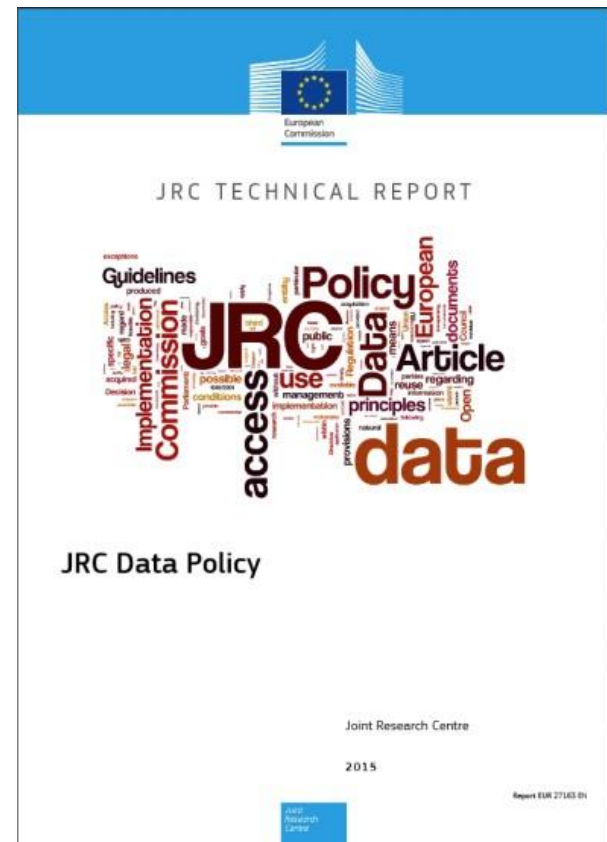
Andrea Perego

SDSVoc 2016

Amsterdam, 30 November 2016

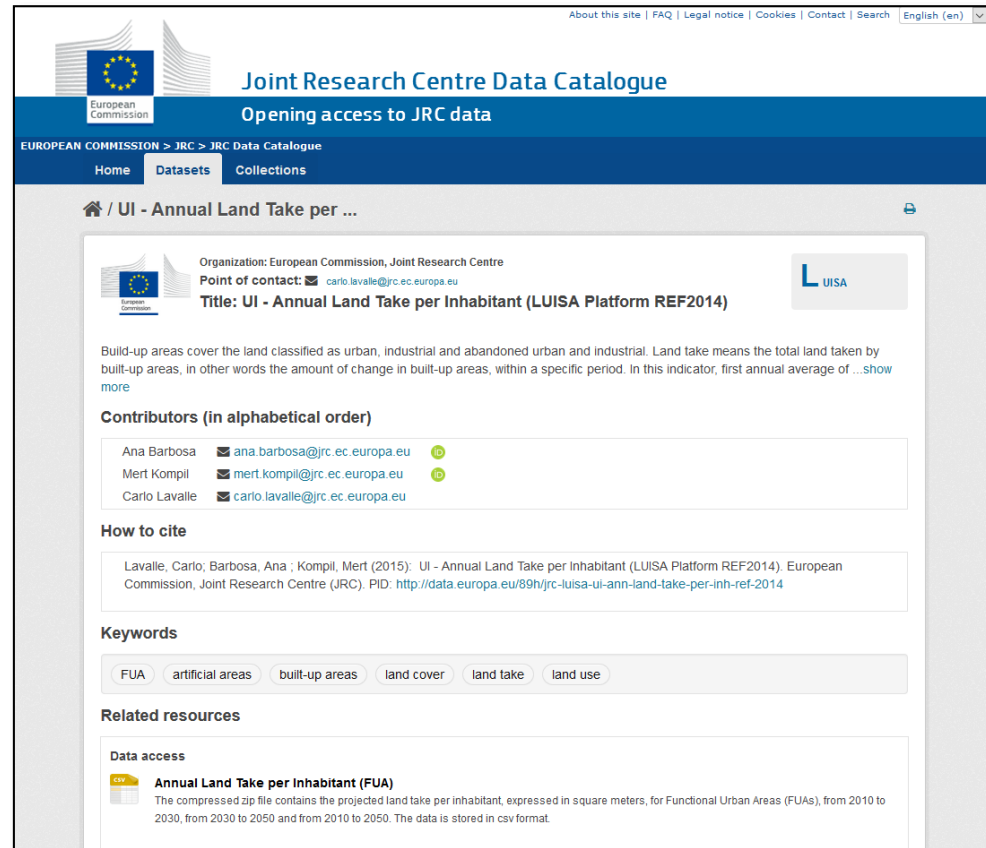
The Joint Research Centre (JRC)

- European Commission's science and knowledge service
- Support to EU policies with independent scientific evidence
- Research covers disciplines concerning all the EU policy areas
- Corporate data policy, based on Open Data principles, adopted in 2015 and now being implemented
- The implementation of the JRC Data Policy includes the setting up of a corporate data catalogue



The JRC Data Catalogue

- Single point of access to all data produced and/or maintained at JRC
- DCAT-AP is the reference metadata standard used, but extended with additional information
- In particular:
 - Metadata elements relevant across scientific domains
 - Metadata elements needed to support data citation



The screenshot shows the JRC Data Catalogue interface. At the top, there is a navigation bar with the European Commission logo and the text 'Joint Research Centre Data Catalogue'. Below this, a blue banner reads 'Opening access to JRC data'. The main content area displays the title 'UI - Annual Land Take per Inhabitant (LUISA Platform REF2014)' and provides contact information for Carlo Lavallo. It also includes a description of the data, a list of contributors (Ana Barbosa, Mert Kompil, Carlo Lavallo), a 'How to cite' section with a citation string, and a 'Data access' section with a CSV file download option.

Organization: European Commission, Joint Research Centre
Point of contact: ✉ carlo.lavalle@jrc.ec.europa.eu
Title: UI - Annual Land Take per Inhabitant (LUISA Platform REF2014)

Build-up areas cover the land classified as urban, industrial and abandoned urban and industrial. Land take means the total land taken by built-up areas, in other words the amount of change in built-up areas, within a specific period. In this indicator, first annual average of ...[show more](#)

Contributors (in alphabetical order)

Ana Barbosa	✉ ana.barbosa@jrc.ec.europa.eu	🟢
Mert Kompil	✉ mert.kompil@jrc.ec.europa.eu	🟢
Carlo Lavallo	✉ carlo.lavalle@jrc.ec.europa.eu	

How to cite

Lavalle, Carlo; Barbosa, Ana; Kompil, Mert (2015): UI - Annual Land Take per Inhabitant (LUISA Platform REF2014). European Commission, Joint Research Centre (JRC). PID: <http://data.europa.eu/89h/jrc-luisa-ui-ann-land-take-per-inh-ref-2014>

Keywords

FUA artificial areas built-up areas land cover land take land use

Related resources

Data access

Annual Land Take per Inhabitant (FUA)
The compressed zip file contains the projected land take per inhabitant, expressed in square meters, for Functional Urban Areas (FUAs), from 2010 to 2030, from 2030 to 2050 and from 2010 to 2050. The data is stored in csv format.

Information relevant for scientific data

Metadata	DCAT-AP	GeoDCAT-AP	DCAT-AP-JRC
Dataset authors		dct:creator	
Data lineage	dct:provenance		
How to use the data			vann:usageNote
Scientific publications			dct:isReferencedBy
Input data	dct:source		

- Only 2 additional properties added wrt what supported in Geo/DCAT-AP
- For more sophisticated specification of data lineage (i.e., machine-readable instead of free-text) we are considering the use of the PROV ontology
- In such a case, lineage includes all the entities involved in the production of the dataset – workflow included

Data citation

- “Making data citable as done with scientific publications”
- Gaining more and more importance in the scientific community
- Based on the use of persistent identifiers (typically, DOIs for data, ORCIDs for data authors – when available)
- Potentially relevant also outside the scientific domain – advantages:
 - Data persistence
 - Data reproducibility
- Reference metadata standard: **DataCite**
- Can DCAT-AP be used for this?

DataCite: Mandatory elements

DataCite 4.0	DCAT-AP 1.1	Comments
Identifier	<i>Partially</i>	DataCite requires this to be a DOI, whereas DCAT-AP does not have such requirement
Creator	No	This agent role is supported in GeoDCAT-AP
Title	Yes	
Publisher	Yes	
Publication year	Yes	

DataCite: Recommended elements

DataCite 4.0	DCAT-AP 1.1	Comments
Subject	Yes	
Contributor	<i>Partially</i>	Geo/DCAT-AP supports only 2 out of the 21 DataCite contributor types
Date	<i>Partially</i>	Geo/DCAT-AP supports only 3 out of the 9 DataCite date types
Resource type	<i>Partially</i>	Geo/DCAT-AP supports only 2 out of the 14 DataCite resource types
Related identifier	Yes	
Description	Yes	
Geolocation	Yes	

DataCite: Optional elements

DataCite 4.0	DCAT-AP 1.1	Comments
Language	Yes	
Alternate identifier	Yes	
Size	Yes	In DCAT-AP, this is a property of the dataset distribution, and not of the dataset itself
Format	Yes	Same as above
Version	Yes	
Rights	Yes	<ul style="list-style-type: none"> • DataCite does not use specific elements for use conditions (i.e., licences) and access rights • In DCAT-AP, use conditions are a property of the dataset distribution, whereas access rights are associated with the dataset
Funding Reference	No	

Persistent identifiers

- Widely used in the scientific community, especially for publications, but now increasingly for authors and data
- Different approaches are used for representing them in RDF – best practices are needed to enable their effective use across platforms
- But more importantly: How can we make them actionable, irrespective of the platforms they are used in?



arXiv.org



Identifiers, by using DCAT-AP

Encoding	Primary ID	Alternative ID	Type (DOI, ORCID, etc.)
As an HTTP URI	@rdf:about	owl:sameAs	-
As a literal	dct:identifier	adms:identifier	@rdf:datatype

- Encoding identifiers as HTTP URIs seems to be the most effective way of making them actionable
- Quite a few identifier schemes can be encoded as dereferenceable HTTP URIs, and some of them are also returning machine readable metadata (e.g., DOIs, ORCIDs)
- They can still be encoded as literals, especially if there is the need of knowing the identifier “type”
- In such a case, a common identifier type registry would ensure interoperability

Agent roles

- Each metadata standard has their own. E.g.:
 - ISO 19115: 20 roles
 - DataCite: 20+ roles
 - Geo/DCAT-AP: 4 roles
- They all use their own vocabularies / code lists
- Two main issues:
 - How to ensure interoperability?
 - Does it make sense to support all of them across platforms?

owner creator
distributor supervisor **user**
principal investigator
curator publisher
sponsor **originator**
producer
contact point
project manager
processor
project leader
editor custodian
funder
registration agency
provider

Agent roles in research data

- Important to denote the type of contribution provided by each individual / organization for producing data
- In some cases, an additional requirement is to specify the temporal dimension of a role – i.e., the time frame during which an individual / organisation played a given role
- And, maybe, also other information – e.g., the organisation where the individual held a given position while playing that role
- The PROV ontology could be used for this purpose, to specify a “qualified attribution”
- To address the different use cases, such “qualified roles” should however be compatible with the corresponding non-qualified forms, and both should be mutually inferable

Agent roles in GeoDCAT-AP

Qualified form

```
a:Dataset a dcat:Dataset;  
  prov:qualifiedAttribution [ a prov:Attribution ;  
# The agent role, as per ISO 19115  
  dct:type <http://inspire.ec.europa.eu/metadata-  
codeList/ResponsiblePartyRole/owner> ;  
# The agent playing that role  
  prov:agent [ a foaf:Organization ;  
    foaf:name "European Union"@en ] ] .
```

Non-qualified form

```
a:Dataset a dcat:Dataset;  
  dct:rightsHolder [ a foaf:Organization ;  
    foaf:name "European Union"@en ] .
```

Publishing metadata on the Web

Why doing this properly?

- Increase data visibility - as well as of the catalogues giving access to them
- Simplifying data discovery from the end users' side

How?

- Embedding metadata in Web pages by using mechanisms as HTML+RDFa, Microdata, Microformats, JSON-LD
- Using fit-for-purpose vocabularies, as Schema.org

Mapping DCAT-AP to Schema.org

- Schema.org includes a number of gaps, preventing the full mapping of DCAT-AP. They include:
 - Identifiers – actually, this is being worked upon:
<http://webschemas.org/identifier>
 - Categories and category schemes
- Is this a problem? DCAT-AP and Schema.org address different use cases
- Rather, the issue is identifying what is important to be mapped – i.e., what information it is useful to be indexed to improve data discovery
- About the use of Schema.org for research data:
 - <https://developers.google.com/search/docs/data-types/datasets>



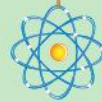
European
Commission

Thanks for your attention!

andrea.perego@jrc.ec.europa.eu

Joint Research Centre

the European Commission's
in-house science service



Joint
Research
Centre



For more information

- JRC Data Policy (doi:[10.2788/607378](https://doi.org/10.2788/607378))
<https://ec.europa.eu/jrc/en/about/jrc-in-brief/data-policy>
- JRC Data Catalogue
<http://data.jrc.ec.europa.eu/>
- DataCite to DCAT-AP Mapping
<https://webgate.ec.europa.eu/CITnet/stash/projects/ODCKAN/repos/datacite-to-dcat-ap/>
- DCAT-AP to Schema.org Mapping
<https://webgate.ec.europa.eu/CITnet/stash/projects/ODCKAN/repos/dcat-ap-to-schema.org/>