

Documenting metadata of language resources in Datahub

This card explains how to add some basic metadata to resources documented in Datahub

Target audience and Scope

The card is intended for people who have a dataset in linked data and wish to document it in Datahub. This card extends the card “*Contributing to the Linguistic Linked Open Data Cloud*” by adding more details on how to document some basic metadata. See also, as an alternative, the “Dataset description with DataID” card.

Registering with Datahub

Go to <http://datahub.io> and follow the steps in the “*Contributing to the Linguistic Linked Open Data Cloud*” reference card. As result, Datahub automatically creates an RDF metadata file for the dataset, based on the Data Catalogue Vocabulary (DCAT). This metadata file is accessible at

http://datahub.io/dataset/DATASET_ID.rdf

where DATASET_ID is the identifier assigned by Datahub to the resource. For example: <http://datahub.io/dataset/apertium-en-es.rdf>

Datahub identifies the dataset by assigning a URI to it with the form:

http://datahub.io/dataset/DATASET_ID

As for instance <http://datahub.io/dataset/apertium-en-es>

The metadata created by Datahub contains the information introduced by hand by the dataset publisher. This information, although useful, has some limitations. For instance the language(s) of the resource, its provenance, or its license are not properly represented. Also, it could be interesting to add other related information.

The default metadata generated by Datahub can be extended with such missing information (provenance, licensing, language, related resources, etc.). To that end, a new RDF file should be created by the dataset publisher with the new metadata, and its information linked to the metadata generated by Datahub.

The extended metadata file should be published as well as part of the dataset in Datahub.

Extending the metadata

First, the data publisher should assign a URI to the dataset. Then, an `owl:sameAs` clause has to be included to express that such URI and the URI created by Datahub refer to the same thing.

Provenance

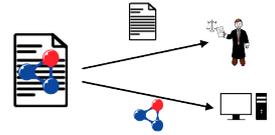
Use the Dublin Core vocabulary for representing provenance (`dct:source`), or PROV-O if you need a richer representation mechanism.

License

A link should be provided to the license by means of `dct:license`. Ideally the license should be represented via a dereferenceable URI associated with an RDF description for machines and an html description for humans.

This page describes in RDF the most popular licenses: <http://rdflicense.appspot.com/>.

See the “*Licensing Language Resources*” reference card for advanced licenses declaration.



Language

The language(s) of the resource can be represented by using `dct:language`. The value of the language property should be a URI representation of the language as in <http://id.loc.gov/> or in <http://www.lexvo.org/>

Related resources

Any additional related information can be linked to the resource by means of other links such as `rdfs:seeAlso`, or `owl:sameAs` when the same item has been described in other metadata sets (e.g, LREMap).

EXAMPLE

```
<http://linguistic.linkeddata.es/set/apertium/EN-ES> a dcat:Dataset ;
  owl:sameAs <http://datahub.io/dataset/apertium-en-es> ;
  dct:source <http://hdl.handle.net/10230/17110> ;
  dct:license <http://purl.org/NET/rdflicense/gpl3.0> ;
  dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
  dct:language <http://id.loc.gov/vocabulary/iso639-1/es> ;
  rdfs:seeAlso <http://dbpedia.org/resource/Apertium> .
```

Apertium RDF EN-ES



RDF version of the Apertium bilingual dictionary EN-ES. The original dataset (in Lemn) comes from <http://hdl.handle.net/10230/17110>. The RDF version of the lexica is modelled using lemon (<http://lemon-model.net/>) and the translation module (<http://purl.org/net/translation>). More about the Apertium RDF dictionaries at <http://linguistic.linkeddata.es/apertium/>

Data and Resources

	English lexicon URI This online lexicon not only contains the EN lexical entries	More information	Go to resource
	Spanish lexicon URI This online lexicon not only contains the ES lexical entries	More information	Go to resource
	EN-ES translation set URI No description for this resource	More information	Go to resource
	SPARQL endpoint No description for this resource	More information	Go to resource
	Zipped Dump It contains two lexicons (EN, ES) and the Translation Set EN-ES, as well as	Preview	Download
	Dataset description in DCAT This extends the metadata in http://datahub.io/dataset/apertium-en-es	More information	Go to resource

Created on 28/05/2015
Last modified on 18/06/2015