

DCAT & metadata file

KDI 2020

20/11/2020

DCAT - ABSTRACT

DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.

This document defines the schema and provides examples for its use.

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. This can increase the discoverability of datasets and data services. It also makes it possible to have a decentralized approach to publishing data catalogs and makes federated search for datasets across catalogs in multiple sites possible using the same query mechanism and structure.

Aggregated DCAT metadata can serve as a manifest file as part of the digital preservation process.

The namespace for DCAT terms is <http://www.w3.org/ns/dcat#>

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

METADATA (CODEBOOK)

For each project («KDIyear-project name».kdi)

- Creator (Creator = project name)
- Creation date (CreationDate= date)
- KDI edition (KDIYear= 2020)
- Members (Members = role:name1, ...)
 - [where role := PM or KE or DS, multiple occurrences of names needed]
- Codebook (Codebook = «projectName».cd)
- Demo (Demo = «projectName».d)
- Demo Slides (DemoSlides = «projectName».ds)
- Project Slides (ProjectSlides = «projectName».ps)
- Repository (Repository = repository-URL)

METADATA (CODEBOOK)

For each input file («project name»-n.in) [where n := number of input files]

- Creator
- Creation date
- Dataset name
- Dataset description (Short textual description)
- List of column names
 - ColumnNames = (input name, SKG name), ...
- Source data set URL (SourceDatasetURL = ...)
- Reference SKG metadata file (SKG = «project name».skg)
- Reference DKG metadata file (DKG = «project name».dkg)
- Size (number of items) (NumberItems = ...)

METADATA (CODEBOOK)

For SKG («project name».skg)

- Creator (project/ group name)
- Creation date
- Reference LKG metadata file (LKG = «project name».lkg)
- Reference DKG metadata file (DKG = «project name».dkg)
- List of entity types (as from SKG)
 - EntityTypes= e1, ...,en
- List of object properties per entity type (as from SKG)
 - ObjectProperties (e1) = o1, ..., on
 - ...
- List of data properties per entity type (as from SKG)
 - DataProperties (e1) = d1, ..., dn
 - ...

METADATA (CODEBOOK)

For LKG («project name».lkg)

- Creator (project/ group name)
- Creation date
- Reference SKG metadata file (SKG = «project name».skg)
- Reference DKG metadata file (DKG = «project name».dkg)
- Languages (Language = l1, ...,)
- List of words added to UKC (as from Excel Import File(s))
 - Words(l1) = w1, ... , wn
 - ...

METADATA (CODEBOOK)

For DKG («project name».dkg)

- Creator (project/ group name)
- Creation date
- List of input files (Input files = «project name»-1.in, ..., «project name»-n.in)
- Reference SKG metadata file (SKG = «project name».skg)
- Reference LKG metadata file (LKG = «project name».lkg)
- Reference project main metadata file (Project = «KDIyear-project name».kdi)
- Number of entities per entity type
 - $\text{NumberEntities}(e1) = N1$
 - ...

KDI 2020 - Metadata Requirements

Set of Metadata machine readable (JSON) files, including:

- «KDIyear-project name».kdi
- «project name»-n.in (for each input file)
- «project name».skg
- «project name».lkg
- «project name».dkg

Note: In the main project documentation (overleaf doc.) the metadata will be indicated only through the references to this set of JSON resources.

DCAT & metadata file

KDI 2020

20/11/2020