

six barriers for open public data

- 1.lack of information that certain data actually exists and is available
- 2.a lack of clarity of which public authority holds the data
- 3.a lack of clarity about the terms of re-use
- 4.data which is made available only in formats that are difficult or expensive to use
- 5.complicated licensing procedures or prohibitive fees
- 6.exclusive re-use agreements with one commercial actor or re-use restricted to a government-owned company

publishing data with DataId

- DataId provides unambiguous and precise meta-information describing your dataset
- your dataset (metadata) gains an online presence (dataid.dbpedia.org & datahub.io)
- your data becomes searchable (search engines & specialized dataset search on dataid.dbpedia.org)
- providing a simple version control system for your data
- automated discovery of similar datasets

benefits when publishing LOD data

- automated link discovery and link count between datasets
- become a (visual) member of the LOD cloud
- validate your dataset against a given ontology (future feature: powered by RDFUnit)
- In general: keeping your data at 5 ★

in use

1. generate a Dataid

List of datasets

Dataset Title

Main Dataset

Options

+

Legend:

Main Dataset

dataid:version

void:subset

dataid:latestVersion

Generate

Upload to Datahub

Dataset: Main Dataset

* Required Field

General Information

Dataset/Subset URI *

e.g. <http://dbpedia.org/dataid.ttl>

Title *

dct:title

Main Dataset

Label *

rdfs:label

Homepage *

dcat:landingPage

Example Resource *

void:exampleResource

Language *

dct:language

Root Resource

void:rootResource

Ontology Location *

dataid:ontologyLocation

Description and rights

Issued date *

dct:issued

9-4-2015

Version Info *

dataid:versionInfo

Contact Information:

Contact Label	Contact Property	Edit	Remove
(Contact Information list is empty)			

Add Contact

Provenance:

WasDerivedFrom

?

Title

dct:title

Contact Label	Contact Property	Edit	Remove
(Contact Information list is empty)			

Add Contact

WasGeneratedBy

?


Title

dct:title

Contact Label	Contact Property	Edit	Remove
(Contact Information list is empty)			

Add Contact

2. publish and revise any time



The easy way to get, use and share data

DatasetsOrganizationsAboutBlogHelp


Search

Home / Organizations / AKSW / **Brown Corpus in RDF/NIF**

Brown Corpus in RDF/NIF

Followers
0

Organization



AKSW
The Research Group Agile Knowledge Engineering and Semantic Web (AKSW) is hosted by the Chair of Business Information Systems (BIS) of the Institute of Computer Science (IfI) /...
[read more](#)

Social

Dataset Activity Stream Related




Brown Corpus in RDF/NIF

RDF version of the Brown Corpus (W. N. Francis, H. Kucera; Brown University; 1979). 1,014,312 words in 500 documents, taken from newspapers texts on diverse topics, non-fiction and fiction books as well as government documents.

Original corpus contains manually annotated sentence and token boundaries as well as word class annotations (such as POS, inflectional morphemes, such as noun plural, verb tense and adjective comparison and special tags for foreign words and proper nouns).

Converted corpus contains complete texts reconstructed from TEI/XML version of the Brown corpus. Word classes were linked via OLiA to ontological categories for aggregated querying.

Data and Resources

	dump folder Contains all 500 converted documents of the original corpus.	More information	Go to resource
	Example corpus document One of the converted corpus files in turtle format	More information	Go to resource
	DataID of the NIF Brown Corpus Turtle file describing the dataset and its distributions	More information	Go to resource

corpus linguistic linguistics lod nif rdf

3. search other datasets

- Cool pic here, not available yet...

Dont know...