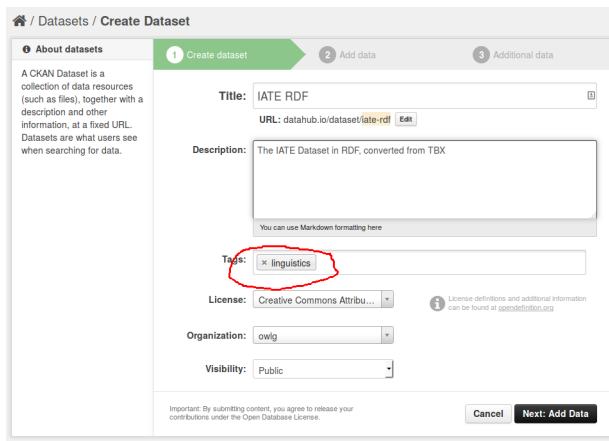


# Contributing to the Linguistic Linked Open Data Cloud

This card describes how you can contribute a resource to the linguistic linked open data cloud by registering it with Datahub. This card is intended for people who have a dataset in linked data and wish to make it visible in the LLOD cloud diagram

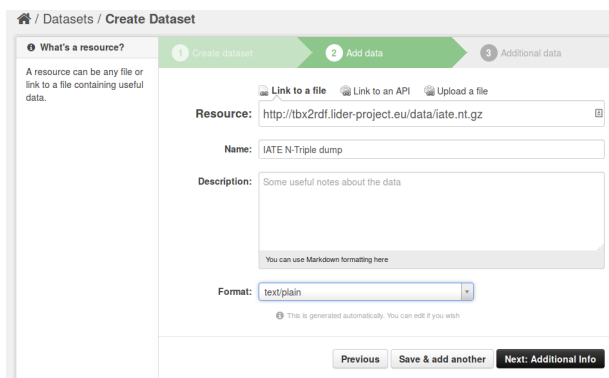
## Registering with Datahub

Go to <http://datahub.io> and choose “Add Dataset”



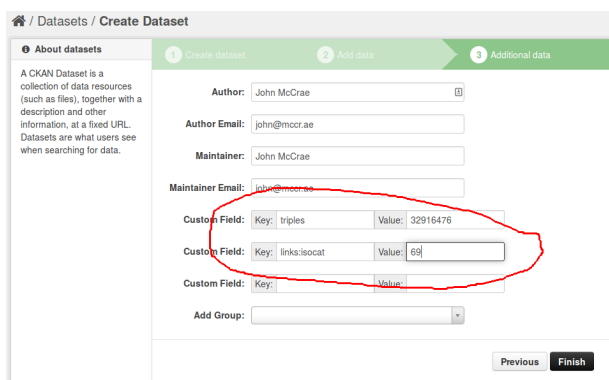
The screenshot shows the 'Create Dataset' form in Datahub. The 'Title' field is 'IATE RDF', the 'URL' is 'datahub.io/dataset/iate-rdf', and the 'Description' is 'The IATE Dataset in RDF, converted from TBX'. The 'Tags' field is circled in red and contains the tag 'linguistics'. The 'License' is 'Creative Commons Attribution', the 'Organization' is 'owlg', and the 'Visibility' is 'Public'.

Give title, description and license and tag as *linguistics* and one of *corpus*, *lexicon*, *language\_description*



The screenshot shows the 'Add data' step of the 'Create Dataset' form. The 'Resource' field is 'http://tbx2rdf.lider-project.eu/data/iate.nt.gz', the 'Name' is 'IATE N-Triple dump', and the 'Description' is 'Some useful notes about the data'. The 'Format' is 'text/plain'.

Add two custom fields giving the size of your resource and the links to other datasets. The key for links is `links:xxx` where xxx is the dataset name in Datahub



The screenshot shows the 'Additional data' step of the 'Create Dataset' form. The 'Author' is 'John McCrae', the 'Author Email' is 'john@mccr.ae', and the 'Maintainer' is 'John McCrae'. The 'Custom Fields' section is circled in red and contains two fields: 'triples' with value '32916476' and 'links:isocat' with value '69'.

# Requirements

## Resolvable

*There must be resolvable <http://> (or <https://>) URIs.*

You must be able to provide a URL giving at least one example e.g., <http://www.myproject.com/example.rdf>

## Links

*The dataset must be connected via RDF links to a dataset that is already in the diagram. This means, either your dataset must use URIs from the other dataset, or vice versa. We require at least 50 links.*

## Accessible

*Access of the entire dataset must be possible via RDF crawling, via an RDF dump, or via a SPARQL endpoint.*

This may be provided by either ensuring

1. There is an "index" page with links to every part of the resource
2. The result is downloadable as a single file or a zip of files
3. All the data is loaded into a RDF store and is queryable using SPARQL

## RDF

*They must resolve, with or without content negotiation, to RDF data in one of the popular RDF formats (RDFa, RDF/XML, Turtle, N-Triples)*

The example URL must return a valid RDF document in RDF/XML to the following query

```
curl -L -H "Accept: application/rdf+xml" \  
http://www.myproject.com/example.rdf
```

## 1000 Triples

*The dataset must contain at least 1000 triples.*

If your dataset is greater than 1MB in size it will likely fit this criterion.

## Linguistic

*Your data must be a language resource*

We define language resources as follows:

Language resources include language data and descriptions in machine readable form used to assist and augment language processing applications, such as written or spoken corpora and lexica, multimodal resources, grammars, terminology or domain specific databases and dictionaries, ontologies, multimedia databases