

Automatischer Upload von Metadaten in data repository

- Datenrepositories wie <http://datahub.io> dienen der besseren Organisation von Datensätzen
- Metadaten zu den Datensätzen, wie Autoren, Größe des Datensatzes, Links zu anderen Datensätzen ebenfalls im Repository

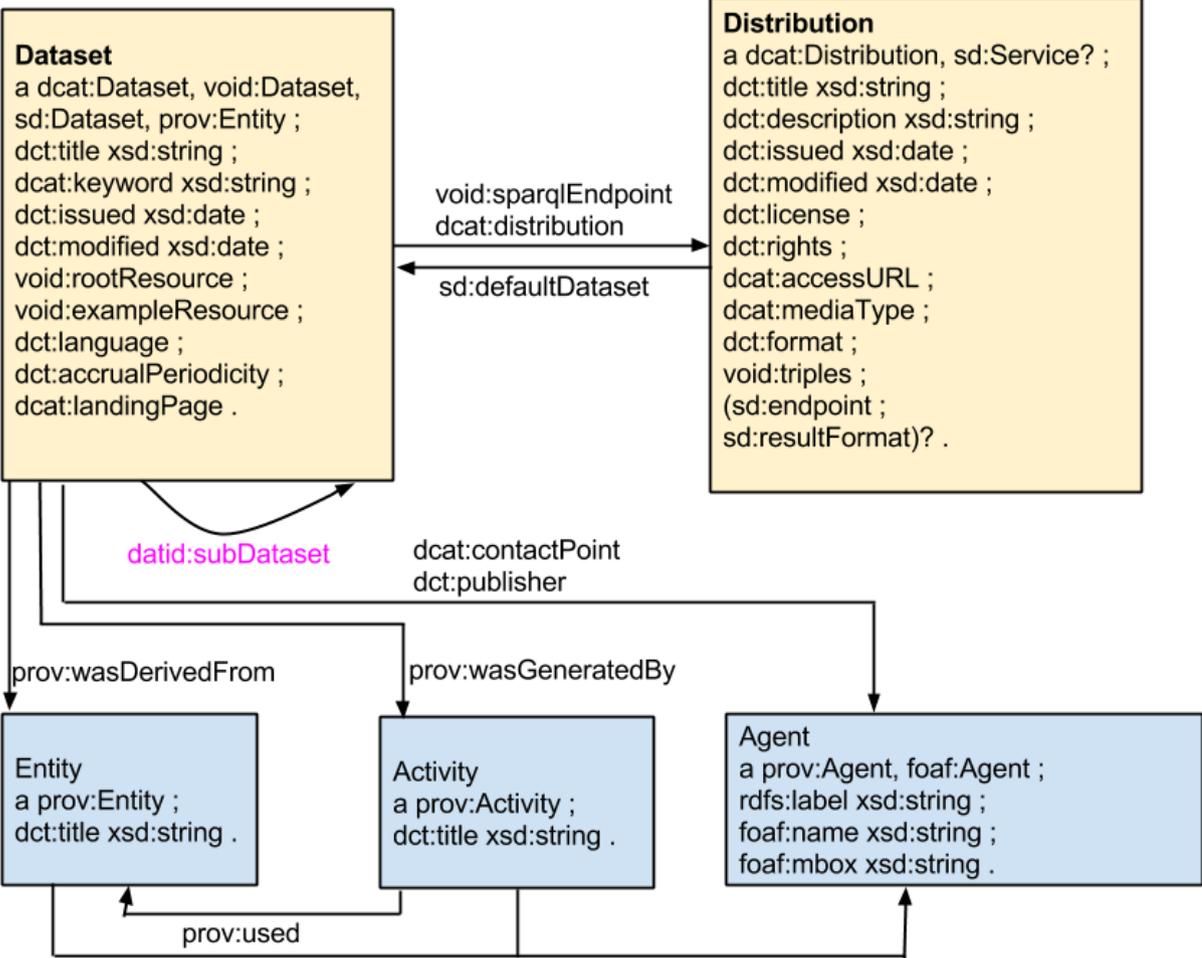
Automatischer Upload von Metadaten in data repository

- Wartung dadurch aufwendig: Bei Änderung des Datensatzes müssen Metadaten im Repository angepasst werden
- Besser: Datensatz liefert selbst Beschreibung seiner Metadaten: **DataID**

DataID

- RDF Datei, die strukturierte Metadatenbeschreibung enthält
- Gleiche Struktur für verschiedene Datensätze: Maschinenlesbare Dokumentation der Daten
- wird aktuell von DBpedia entwickelt:
<http://wiki.dbpedia.org/coop/DataIDUnit>

DataID



@prefix dct: <http://purl.org/dc/terms/> .
 @prefix dcat: <http://www.w3.org/ns/dcat#> .
 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
 @prefix sd: <http://www.w3.org/ns/sparql-service-description#> .
 @prefix void: <http://rdfs.org/ns/void#> .
 @prefix datid: <????> .

dct:creator
 dct:publisher
 dct:contributor

proposed property,
 not yet established

Datahub.io

- Derzeit eines der bedeutendsten Repositories
- basiert auf CKAN framework
- Upload von Metadaten möglich via Python API (<https://github.com/ckan/ckanapi>)
- kein strukturiertes Metadatenmodell, sondern Key – value Paare und Ressourcen Dateien

Tasks

- Implementierung eines Tools, das Metadaten aus DataID-Dateien extrahiert, Datahub-Kompatibel umformt und hochlädt
- CLI ausreichend, GUI nice to have

Anforderungen

- Python
- OWL