

DELIVERABLE D1: INTERNAL LINKS ESTABLISHED

Author:
Victor de Boer (VU Amsterdam)

Deliverable type: DATA

Version	Date	Author	Description
0.1	26-03-2014	Victor de Boer	First version
0.2	16-04-2014	Victor de Boer	Final version

This deliverable describes the internal links established in the DSS project.

OVERVIEW

The figure below gives a global overview of the internal and external links. The links take form of RDF triples of which the subject and object are in separate named graphs (datasets). We here list the internal links. In Deliverable D7, we list the external links.

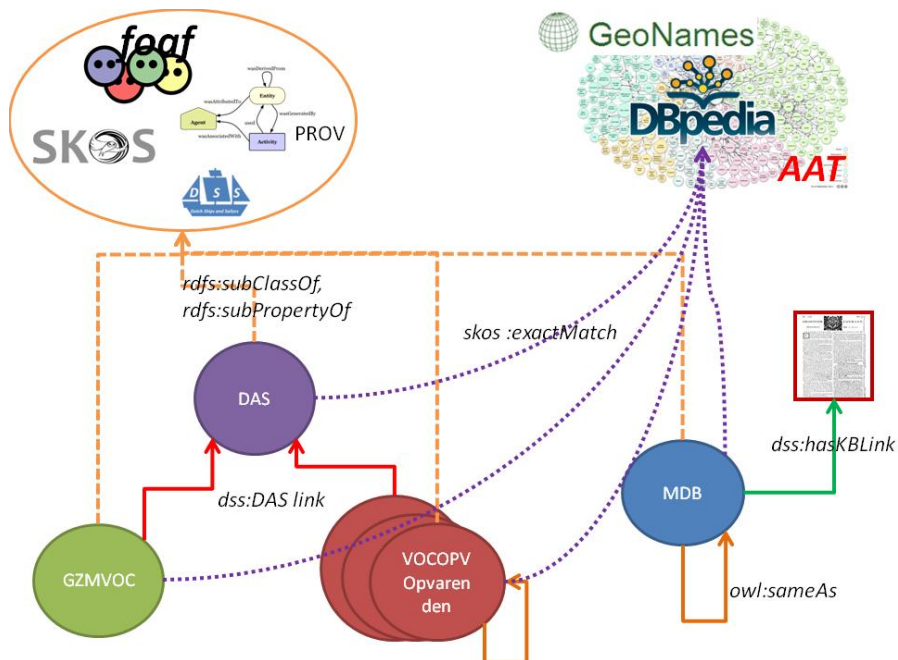


FIG 1: INTERNAL AND EXTERLAN LINKS IN THE DSS DATA CLOUD

GZMVOC TO DAS

The Generale Zeemonsterrollen dataset uses explicit codes references to DAS voyages. These were used to generate explicit links between GZMVOC and DAS. We use two different RDF properties, which correspond to the original metadata fields. All links are stored in a separate named graph (http://purl.org/collections/nl/dss/gzmvoc/gzmvoc_2_das.ttl). The table below lists the number of links. In total 5303 links are established.

Predicate	#Triples	#Distinct subjects	#Distinct objects
dss:gzmvoc/has_das_link_heen	3607	3607	1665
dss:gzmvoc/has_das_link_terug	1696	1696	1134

VOC OPVARENDEN TO DAS

The original VOC Opvarenden dataset uses explicit codes references to DAS voyages. These were used to generate explicit links between VOC Opvarenden and DAS. We use three different RDF properties, which correspond to the original metadata fields. All links are stored in a separate named graph (http://purl.org/collections/nl/dss/vocopv/vocopv_2_das.ttl.gz). The table below lists the internal links. In total 1,128,416 links are established.

Predicate	#Triples	#Distinct subjects	#Distinct objects	Domain(s)	Range(s)
dss:vocopv/has_dasuitreis	767581	767581	3135	rdfs:Resource	rdfs:Resource
dss:vocopv/has_tblSchipKaapDASURI	45350	45350	3182	rdfs:Resource	rdfs:Resource
dss:vocopv/has_vwPersoonDASURI	315485	314932	5832	rdfs:Resource	rdfs:Resource

MDB SAMEAS LINKS

In the MDB dataset many ships occur multiple times, however it is initially unknown which ships are which. We therefore assume that all ships are unique and only at a later state attempt to identify recurring ships. This is done using different algorithms, designed and implemented within the context of the Master thesis of VU student Robin Ponstein. A sample of his results were evaluate by Jur Leinenga and an acceptable precision was found. Still, the links are stored in a separate named graph

(http://purl.org/collections/nl/dss/mdb/mdb_ship_sameas.ttl) with appropriate provenance and content confidence metadata. The table below lists the sameas links currently in the Huygens data store

Predicate	#Triples	#Distinct subjects	#Distinct objects
owl:sameAs	33435	7291	7291

SCHEMA LINKS

A number of dataset-specific properties have been mapped to a common DSS schema, which essentially establish internal links as well. We do not list the numbers but refer the reader to deliverable D2.