CLARIN-DSS "Dutch Ships and Sailors"

DELIVERABLE D3: DATASETS CONVERTED

Author:

Victor de Boer (VU Amsterdam)

Deliverable type: DATA

Version	Date	Author	Description
0.1	04-10-2013	Victor de Boer	Datasets converted
0.2	04-03-2014	Victor de Boer	First version of deliverable
1.0	19-04-2014	Victor de Boer	Final version

More details can be found in Deliverable D10. This deliverable describes the results of the conversion of the four datasets to RDF. The converted datasets are available as RDF named graphs in a number of forms:

- Through the semantic layer http://dutchshipsandsailors.nl/dss, or at the backup server at http://semanticweb.cs.vu.nl/dss The graphs can be browsed there and downloaded as RDF/Turtle or RDF/XML.
- Through the public Github repository https://www.github.com/biktorrr/dss. Here up to date RDF/Turtle versions of all datasets are downloadable separately or in bulk.
- Through DANS EASY (to be specified)

We list the resulting conversions of the four datasets in separate sections. In the table below we present the overall statistics for all named graphs (datasets) produced. The up-to-date version of this table can be found at http://www.dutchshipsandsailors.nl/data/browse/list_graphs . In this table we list the number of RDF triples as well as which of the datasets the named graph belongs to.

RDF Graph	Triples	Dataset
http://purl.org/collections/nl/dss/vocopv/vocopv_opv.ttl.gz	19,104,514	VOC Opvarenden
http://purl.org/collections/nl/dss/vocopv/vocopv_sol.ttl.gz	2,231,367	VOC Opvarenden
http://purl.org/collections/nl/dss/mdb/mdb_data.ttl.gz	1,296,641	MDB
http://purl.org/collections/nl/dss/vocopv/vocopv_2_das.ttl.gz	1,128,416	VOC Opvarenden
http://purl.org/collections/nl/dss/vocopv/vocopv_beg.ttl.gz	636,333	VOC Opvarenden
http://sws.geonames.org/geonames-NL.ttl	309,678	External
http://e-culture.multimedian.nl/ns/rkd/aatned/aatned.rdf	264,968	External
http://purl.org/collections/nl/dss/mdb/mdb_2_kb.ttl	179,120	MDB
http://purl.org/collections/nl/dss/das/das_data.ttl	149,357	DAS
http://purl.org/collections/nl/dss/gzmvoc/gzmvoc_data.ttl	110,986	GZMVOC

http://sws.geonames.org/geonames_nl_as_skos.ttl	42,811	External
http://purl.org/collections/nl/dss/mdb/mdb_ship_sameas.ttl	33,435	MDB
http://purl.org/collections/nl/dss/vocopv/vocopv_gen_thes.ttl	12,851	VOC Opvarenden
http://purl.org/collections/nl/dss/das/das_thes_gen.ttl	7,034	DAS
http://purl.org/collections/nl/dss/dbpedia_links.ttl	5,449	External
http://purl.org/collections/nl/dss/gzmvoc/gzmvoc_2_das.ttl	5,303	GZMVOC
http://sws.geonames.org/ontology_v2.2.1.rdf	2,895	External
http://purl.org/collections/nl/dss/al_all_place_2_geonames.ttl	2,528	DSS (all)
http://purl.org/collections/nl/dss/mdb/mdb_thes_places.ttl	2,273	MDB
http://purl.org/collections/nl/dss/gzmvoc/gzmvoc_thes_gen_places.ttl	591	GZMVOC
http://purl.org/collections/nl/dss/mdb/mdb_thes_rangen.ttl	585	MDB
http://purl.org/collections/nl/dss/vocopv/vocopv_schema.ttl	337	VOC Opvarenden
http://purl.org/collections/nl/dss/dss_provenance.ttl	273	DSS (all)
http://purl.org/collections/nl/dss/mdb/ranks_and_shiptypes_1.ttl	245	MDB
http://purl.org/collections/nl/dss/gzmvoc/gzmvoc_schema.ttl	232	GZMVOC
http://purl.org/collections/nl/dss/mdb/mdb_thes_generated.ttl	196	MDB
file:///data/cliopatria/ClioPatria/rdf/base/rdfs.rdfs	190	External
http://purl.org/collections/nl/dss/gzmvoc/gzmvoc_thes_gen.ttl	166	GZMVOC
http://purl.org/collections/nl/dss/mdb/mdb_schema.ttl	149	MDB
http://purl.org/collections/nl/dss/das/das_schema.ttl	98	DAS
http://purl.org/collections/nl/dss/dss_schema.ttl	59	DSS (all)
http://e-culture.multimedian.nl/ns/rkd/aatned/aatned.rdfs	27	External
Total #triples:	**25,529,107**	

GENERALE ZEEMONSTEROLLEN (GZMVOC)

The Generale Zeemonsterrollen were converted from the original EXCEL data as provided by Matthias van Rossum to an XML version to allow for XML to RDF conversion. The ClioPatria XMLRDF tool was used to convert the data to an RDF version. The datamodel was based on the original Excel sheet and constructed in collaboration between Victor de Boer and Matthias van Rossum. The figure below shows a simplified datamodel. The RDFS schema file http://purl.org/collections/nl/dss/gzmvoc/gzmvoc_schema.ttl describes the entire datamodel.

The main data named graph (http://purl.org/collections/nl/dss/gzmvoc/gzmvoc_data.ttl) has 110,986 triples, with 58 predicates, 5 RDFS classes.

```
<gzmvoc>
        a void:Dataset ;
        a lib:CloudNode ;
        lib:source gzmvoc: ;
        dc:title "Generale Zeemonsterrollen VOC (Matthias van Rossem)" ;
        dc:description "Generale Zeemonsterrollen VOC (Matthias van Rossem)" ;
        owl:versionInfo "0.1" ;
        void:dataDump
                <gzmvoc_data.ttl> ,
                <gzmvoc_2_das.ttl> ,
                <gzmvoc_schema.ttl>,
                <gzmvoc_thes_gen_places.ttl>,
                <gzmvoc_thes_gen.ttl>.
```
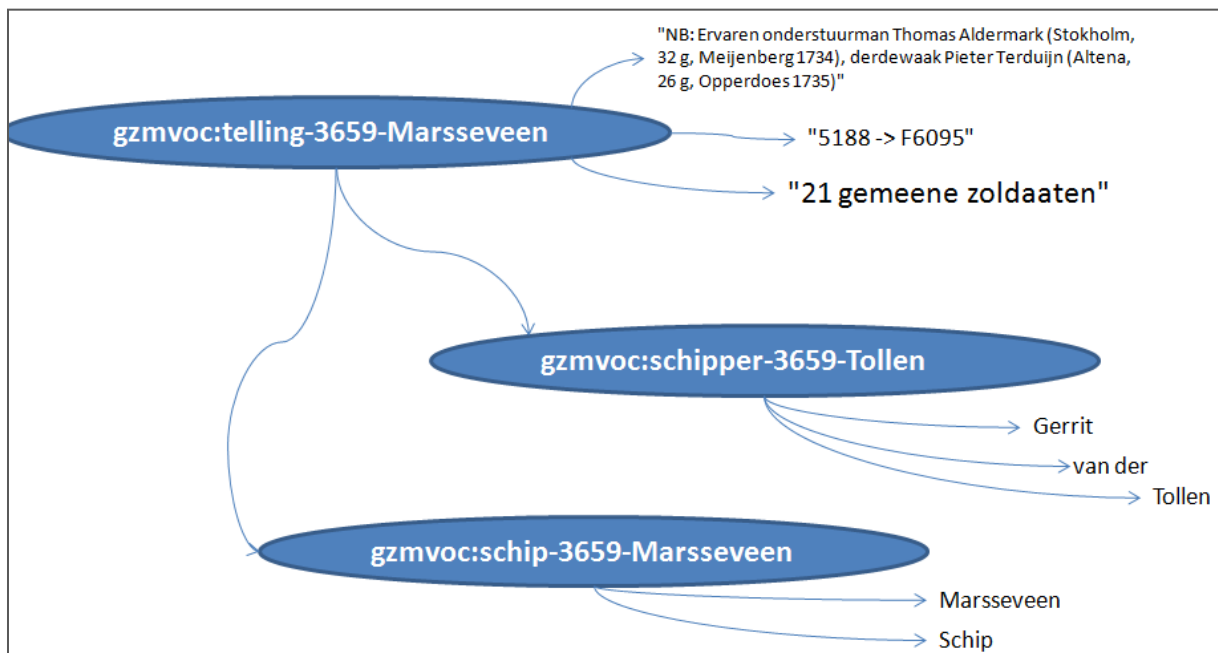
FIG 1: VOID SPECIFICATION FOR GZMVOC DATASET, SHOWING THE RELATED FILES/NAMED GRAPHS

FIG 2: SIMPLIFIED MODEL FOR GZMVOC

NOORDELIJKE MONSTEROLLEN DATABASES (MDB)

The original Noordelijke Monsterollen Databases (MDB) was provided as a SQL dump file (.sql) by the original maker of the data, Jur Leinenga. That data was converted to an XML version by loading it in a XAMPP instance and exporting to that format. The XML version (available on Github) was then converted to RDF with an XMLRDF rewriting script. The model was developed in a collaboration with Jur Leinenga and Victor de Boer, based on the original data model. The simplified datamodel is shown below, the complete RDFS datamodel is found in the named graph http://purl.org/collections/nl/dss/mdb/mdb_schema.ttl. The main data graph http://purl.org/collections/nl/dss/mdb/mdb_data.ttl has 1,296,641 triples, with 27 predicates and 8 classes.

```
<mdb>
        a void:Dataset ;
        a lib:CloudNode ;
        lib:source mdb: ;
        dc:title "Noordelijke Monsterrollen (Jur Leinenga)" ;
        dc:description "Noordelijke Monsterrollen (Jur Leinenga)" ;
        owl:versionInfo "0.1" ;
        void:dataDump
                <mdb_data.ttl> ,
                <mdb_schema.ttl>,
                <mdb_thes_rangen>,
                <mdb_thes_generated>,
                <mdb_thes_places>,
                <alignments/ranks_and_shiptypes_1.ttl>,
                <mdb_2_kb>, # links to KB, by Andrea
                <mdb_ship_sameas>. # sameAs links by Robin.
```
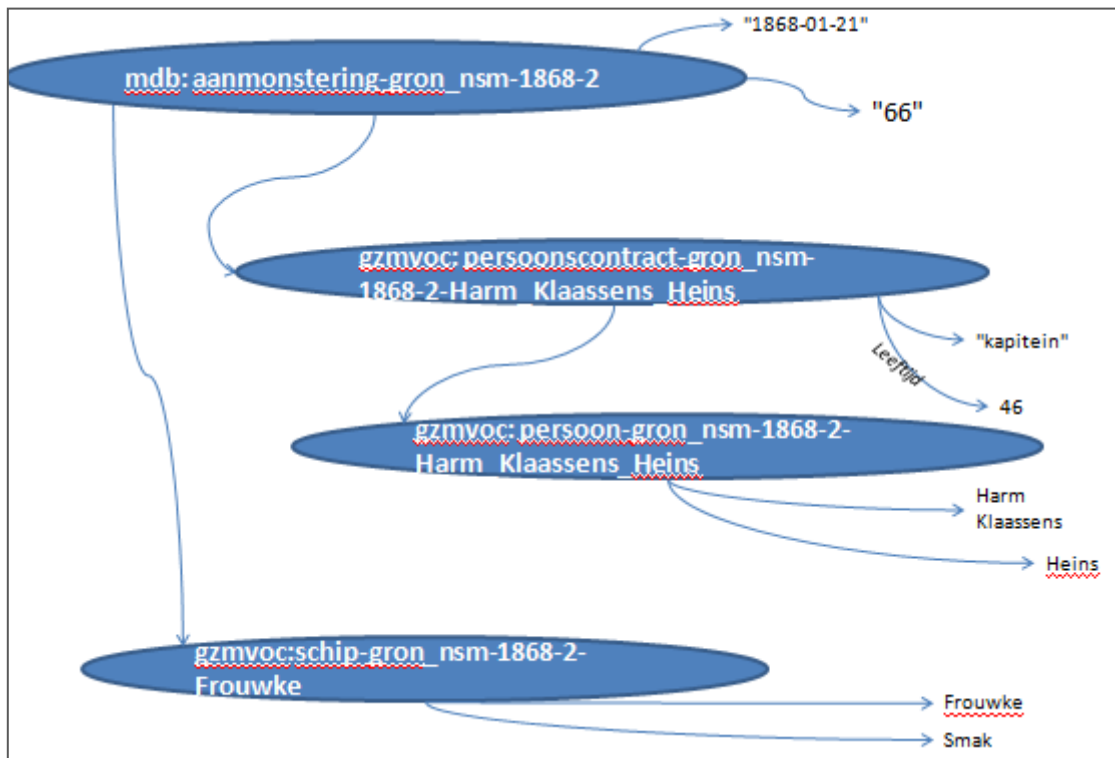
FIG 3: VOID SPECIFICATION FOR MDB DATASET, SHOWING THE RELATED FILES/NAMED GRAPHS

FIG 4: SIMPLIFIED MODEL FOR MDB

VOC Opvarenden (VOCOPV)

The original VOC opvarenden data was downloaded as an CSV file from DANS Easy website (https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:33602). That data was converted to an XML version using a simple python script (available on Github). The XML version was then converted to RDF with an XMLRDF rewriting script. The model was developed in a collaboration with Jur Leinenga, Matthias van Rossum and Victor de Boer, based on the original data model. The complete RDFS datamodel is found in the named graph. There are three main data files *opvarenden (http://purl.org/collections/nl/dss/vocopv/vocopv_opv.ttl)*, *soldijboeken (http://purl.org/collections/nl/dss/vocopv/vocopv_sol.ttl)* and *begunstigden*

(*http://purl.org/collections/nl/dss/vocopv/vocopv_beg.ttl*) with 19,104,514, 2,231,367 and 636,333 triples respectively.

```
<vocopv>
        a void:Dataset ;
        a lib:CloudNode ;
        lib:source vocopv: ;
        dc:title "VOC Opvarenden " ;
        dc:description "VOC Opvarenden" ;
        owl:versionInfo "0.1" ;
        void:dataDump
                <vocopv_opv.ttl.gz> ,
                <vocopv_beg.ttl> ,
                <vocopv_sol.ttl> ,
                <vocopv_gen_thes.ttl>,
                <vocopv_schema.ttl>,
                <vocopv_2_das.ttl>.
```

FIG 5: VOID SPECIFICATION FOR VOCOPV DATASET, SHOWING THE RELATED FILES/NAMED GRAPHS

DUTCH ASIATIC SHIPPING (DAS)

The original Dutch Asiatic Shipping data was downloaded as an CSV file from Huygens ING website. That data was converted to an XML version using a simple python script (available on Github). The XML version (available on Github) was then converted to RDF with an XMLRDF rewriting script. The model was developed in a collaboration with Jur Leinenga, Matthias van Rossum and Victor de Boer, based on the original data model. The complete RDFS datamodel is found in the named graph http://purl.org/collections/nl/dss/das/das_schema.ttl. The main data graph http://purl.org/collections/nl/dss/das/das_data.ttl has 149,357 triples, with 21 predicates and 6 classes.

```
<das>
        a void:Dataset ;
        a lib:CloudNode ;
        lib:source das: ;
        dc:title "Dutch Asiatic Shipping" ;
        dc:description "DAS Dutch Asiatic Shipping (English, with details)" ;
        owl:versionInfo "0.1" ;
        void:dataDump
                <das_data.ttl> ,
                <das_thes_gen.ttl>,
                <das_schema.ttl>.
```

FIG 6: VOID SPECIFICATION FOR DAS DATASET, SHOWING THE RELATED FILES/NAMED GRAPHS