Dutch Ships and Sailors
Linked Data Cloud

Victor de Boer, Matthias van Rossum, Jur Leinenga, Rik Hoekstra

With input from Andrea Bravo Balado and Robin Ponstein

Netherlands Institute for Sound and Vision / VU University Amsterdam
v.de.boer@vu.nl
ISWC2014
The Problem:
((Maritime) historical) data is not integrated

25+ Maritime datasets; Heterogeneous
The solution

Well, Linked Data obviously!
But why Linked Data

- Heterogeneous models, one dataformat
  - Link what can be linked
  - Keep specificity of original data
  - Allow integration at project level (and beyond)

- Links to other sources: re-use knowledge

- Extensible

- Allow multiple levels of semantic enrichment/normalization
  - through Named Graphs
  - Provenance
Dutch Ships and Sailors

“VOC Opvarenden”
Mustering and payroll information (DANS Easy)

Dutch-Asiatic Shipping (DAS) – Voyages (Huygens ING)

KB Delpher

Werkers van de wereld
Globalisering, maritieme arbeidsmarkten en de verhouding tussen Azia ten en Europeenen in dienst van de VOC
Matthias van Rossum
Modeling in collaboration with historians (1)

Jur Leinenga
(Huygens ING)
Muster-rolls
Northern Provinces
1803-1937
Modeling in collaboration with historians (2)

Matthias van Rossum (VU-hist)
Payroll information for European vs Asiatic Sailors (17th / 18th C)
Modelling principles

• Model each dataset as directly as possible
  – Only “syntactical” transformation to RDF
  – No normalization

• Reusability

• Transparency, trust

• Normalize and link in second stage
  – store in separate RDF Named Graphs
Link properties and classes to interoperability layer

mdb:Schip1

mdb:scheepsType

mdb:Kof

das:ShipX

das:typeOfShip

das:Kofship

dss:has_shipType

rdfs:subPropertyOf

rdfs:subPropertyOf
Vocabulary Links

http://semanticweb.cs.vu.nl/amalgame/

Links to DBPedia (Ship types, places, ranks)
Links to Getty AAT (Ship types, ranks)
Links to GeoNames (Places)
Identifying ships

• Identify ships within a dataset using Machine Learning techniques
  – Based on: name, size, type, destinations etc.
  – Background knowledge

• 33,435 owl:sameAs links

– Robin Ponstein
• Use ML to detect links between ships and historical newspaper articles (delpher.nl)
  – Features: ship name, time intervals, captain’s names, ship type, named entities, keywords, background knowledge
  
• 179,120 links
Example
Provenance (PROV-O)

- Individual *named graphs* have provenance information
  - Who made it (people/software?)
  - Based on what source
  - Content confidence

- Matches historical science requirements
ClioPatria Triplestore

- Data live at Huygens Institute for Dutch History
  - [http://dutchshipsandsailors.nl/data](http://dutchshipsandsailors.nl/data)
  - ~30 Million triples

- Dev. Server
  - [http://semanticweb.cs.vu.nl/dss](http://semanticweb.cs.vu.nl/dss)

- Purl.org URIs redirect to live server w/ content negotiation
- SPARQL endpoint
- Web interface
Data analysis and visualisation

Use the textarea below to fire a SPARQL query at the DSS triple store. You can choose to adapt that query if needed before launching it.

Select Query

Select Query

- Find all mcb aanmonsteringen that have a ship and a captain with the last name "Boer"
- Give me all ships (across datasets) with the name "Johanna"
- Find all mcb aanmonsteringen, and list the last name of the captain of the ship
- Find things in DAS and GZMVOC that match the same place in Geonames
- Find things in 3 datasets that match the same place in Geonames and also give me the lat/long
- Places where DAS ships have been
- Linked newspaper articles for MDB brikken heading to RIGA
- Linked newspaper articles for MDB schoeners with captains name "Veldman"
- Links to CEDAR Historical Occupations
- Alle KB geïnkleerde aanmonsteringen met een kapitein met boer in de naam
- Personen met “jans” in de naam, aangemoederd op schip met “kof” in het type
- MDB Aanmonsteringen op subtypes van kustvaarders (AAT)
- MDB Aanmonsteringen op subtypes van kustvaarders (AAT) in 1815
- Alle VOC opvarendenrecords, met een gemachte plaats en bijbehorende provincie
- GZM voor links naar DAS: VOC kamers met aziatische bemanning
Current work: linking original scans
Take home

• Linked Data principles are a great fit to digital history requirements
  – Heterogeneous models/datasets, light-weight reusable integration
  – Multiple levels of normalisation, through separate named graphs
  – SW Provenance matches Historical Provenance

• Watch out when you sail your Schooner into the North Sea
http://dutchshipsandsailors.nl/data

v.de.boer@vu.nl