Transformations for aggregating Linked Open Data

The Local Amsterdam Cultural Heritage Linked Open Data Network

Lukas Koster (Library of the University of Amsterdam)
Ivo Zandhuis (Ivo Zandhuis Research & Consultancy)

SWIB 2018 Bonn
AdamLink Project

AdamNet Foundation:
Amsterdam based Library Collaboration Association

33 member institutions

- public libraries
- higher education libraries
- special libraries
- museums
- archives
- documentation centers
AdamLink Project Goal

Create a linked open data infrastructure for AdamNet member institutions collections on the topic of "Amsterdam"

Data Only!

And some proof of concept apps
Target audiences

- Researchers
- Education
- General public
- Creative industry
Initial Collections

Amsterdam Museum
- Image database

Amsterdam City Archive
- Image database

International Institute of Social History
- Image database
- Interviews

University of Amsterdam
- Digitised maps

Amsterdam Public Library
- Books on topic "Amsterdam"

Photographs, paintings, drawings, posters, maps, etc.

Texts

Catalogue records
Triply Linked Open Data Platform

- Triple Store
- Import Turtle
- Browser
- Tables
- Sparql Endpoints
- API
- Management
- Hosted

https://triply.cc/
Challenge: Distribution vs. Aggregation

Distribution is up-to-date, but

- Often: no ‘live’ LOD at all
- Local differences in ontologies, vocabularies, data types, data quality
- Complicated data joining
- Performance issues
Dutch National Program “NDE”
Digital Heritage Network

- Service portals: Centralised, optionally distributed
- Knowledge Graph: Originally a temporary provision
- Registry: Centralised, optionally distributed
- Network of Terms: Centralised, aggregated, optionally distributed
- Aggregators
- Collection management systems: Distributed
- Terminology management systems: Distributed
Aggregators: changing role

NDE: extension to the model - “service platforms”

“A service platform combines and enriches heritage information and makes it usable in a specific context”
(https://github.com/netwerk-digitaal-erfgoed/high-level-design/blob/master/building-blocks.md#service-portal)

Ruben Verborgh: “Decentralization needs replication for realistic performance”

“...transparent layers in network of nodes...”
(http://slides.verborgh.org/ELAG-2018/#)
Central Platform: Current situation

![Diagram showing the central platform with connections to various datasets and systems.](image-url)
http://data.adamlink.nl/
Local workflows

- ImageBase (HomeGrown)
- XML + MARC
- ImageBase (HomeGrown) + XSLT + Python
- Adlib
- Dublin Core
- Dublin Core + Proprietary
- PHP
- Triply RDF Turtle
- IISG-Triply
- Evergreen
- N-Triples
- SOR (OAIS)
- Python
- schema.org
- ProQuest AquaBrowser
- data.bibliotheken.nl
adamlink.nl: "linking points"

- **Locations**
  - Streets
  - Buildings

- **Creators/contributors**
  - People
  - Organisations

- **Types**

- **Subjects**
  - Including
  - People
  - Organisations
What we need is *transformation*

“Aggregation” into one dataset at one endpoint

- https://r4ds.had.co.nz/introduction.html
Transformations

1. Ontology alignment
2. Authority alignment
3. Object types alignment
4. Additional statements
5. Restructuring data
6. Data typing
Example: Herengracht / Springer, C.
Example: Herengracht / Springer, C.

<http://hdl.handle.net/11259/collection.37414>

rdf:type       edm:ProvidedCHO ;
rdfs:label     "De bocht in de Herengracht"^^xsd:string ;
dc:identifier  "SA 286"^^xsd:string ;
dc:title       "De bocht in de Herengracht"^^xsd:string ;
sem:hasBeginTimeStamp  "1882"^^xsd:string ;
sem:hasEndTimeStamp    "1882"^^xsd:string ;
foaf:depiction  <http://amasp.adlibhosting.com/wwwopacx_images...> ;
dc:creator    <https://adamlink.nl/person/springer-cornelis/2220> ;
dc:subject    "wintergezicht"^^xsd:string ;
dc:type       <http://vocab.getty.edu/aat/300177435> ;
dct:spatial   <https://adamlink.nl/geo/street/herengracht/1768> ;
edm:isShownAt <http://hdl.handle.net/11259/collection.37414> ;
dc:rights     <http://creativecommons.org/publicdomain/mark/1.0/> ;
(1/6) Ontology alignment: Common Data model?

Europeana Data Model?
- Targeted at data aggregation/consolidation
- Focused on internal Europeana workflow
- Mainly DC properties
- EDM properties: internal procedures

Dublin Core?
- Not rich enough

CIDOC-CRM?
- Not simple enough
(1/6) Ontology Alignment: Common properties

- rdf:type
- foaf:depiction
- rdfs:label
- dc:creator
- dc:contributor
- dc:type
- dc:subject
- dct:spatial
- dct:temporal
- dc:date
- sem:hasBeginTimeStamp
- sem:hasEndTimeStamp
(2/6) Authority alignment: Common vocabularies

- **Type**
  - AAT

- **People**
  - VIAF
  - WikiData
  - RKD-Artists
  - ...

- **Subjects**
  - WikiData
  - (spatial) Geonames, TGN
  - ...

---

[Image of a card with a logo: A'DAMNET]
(2/6) Authority alignment: sameAs

- types - AAT URI
- creator/contributor -> AdamLink URI
- subject (for persons) -> AdamLink URI
- location -> AdamLink URI
(3/6) Object types alignment

rdf:type:

- edm:ProvidedCHO
- schema:Person, schema:Organization
- hg:Street, hg:Building, hg:District
(4/6) Additional statements

Data management metadata
organisations, datasets

- void:inDataset - AdamLink dataset URI
- dcterms:publisher - AdamLink Org URI
- dc:type - AAT URI, derived from rdf:type
(5/6) Restructuring data

eg. From EDM to DC:

≡ Rijksmuseum data
≡ complex path (via ore:Aggregation-edm:WebResource) between object-uri and the location of an image
≡ transformed into foaf:depiction
(6/6) Data typing

Complex, not fully developed:
xsd:string, xsd:int, xsd:date

or should this be the responsibility of data provider?
Conclusion

- decentral vs central
- aggregation or not
- transformation

... *all depend on the use case!*

We tried to develop an overview of the steps needed in transformation

Discussion: who is responsible for what?
That's all Folks!

https://tinyurl.com/adamlinkSWIB18