

Amsterdam Museum Linked Open Data

Victor de Boer^{a,*} Jan Wielemaker^a Judith van Gent^b Marijke Oosterbroek^b Michiel Hildebrand^a
Antoine Isaac^a Jacco van Ossenbruggen^a Guus Schreiber^a

^a *Department of Computer Science, VU University Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands*

Email: {v.de.boer, j.wielemaker, a.isaac, m.hildebrand, j.r.van.ossenbruggen, guus.schreiber}@vu.nl

^b *Amsterdam Museum, Postbus 3302, 1001 AC Amsterdam, The Netherlands*

Email: {j.vangent, m.oosterbroek}@amsterdammuseum.nl

Abstract. In this document we describe the Amsterdam Museum Linked Open Data set. The dataset is a five-star Linked Data representation and comprises the entire collection of the Amsterdam Museum consisting of more than 70.000 object descriptions. Furthermore, the institution's thesaurus and person authority files used in the object metadata are included in the Linked Data set. The data is mapped to the Europeana Data Model, utilizing Dublin Core, SKOS, RDA-group2 elements and the OAI-ORE model to represent the museum data. Vocabulary concepts are mapped to Geonames and DBpedia. The two main contributions of this dataset are the inclusion of internal vocabularies and the fact that the complexity of the original dataset is retained.

Keywords: Cultural Heritage, Museum, Thesaurus, Europeana Data Model

1. Introduction

In this document, we describe the Amsterdam Museum Linked Data set. The Amsterdam Museum¹ is a Dutch museum hosting objects related to the history and culture of Amsterdam and its citizens. Among these objects are paintings, drawings, prints, glass and silver objects, furniture, books, costumes, etc. At any given moment, around 20% of the objects are on display in the museum's exhibition rooms, while the rest is stored in storage depots.

As do many museums, the Amsterdam Museum uses a digital data management system to manage their collection metadata and authority files, in this case the proprietary Adlib Museum software². As part of the museum's policy of sharing knowledge, in 2010, the Amsterdam Museum made their entire collection available online using a creative commons license. The collection can be browsed through a web-interface³.

Second, for machine consumption, an XML REST API was provided that can be used to harvest the entire collection's metadata or retrieve specific results based on search-terms such as on creator or year⁴. The latter API has been used extensively in multiple Cultural Heritage-related application-building challenges⁵.

While larger cultural heritage institutions such as the German National Library⁶ or British National Library⁷ have the resources to produce their own Linked Data, metadata from smaller institutions is currently only being added through large-scale aggregators such as Europeana⁸. In current workflow of the Europeana Linked Data pilot⁹, metadata records are ingested and restructured to fit the Europeana Data Model (EDM)[2] and published it as "five-star" Linked Data on Europeana servers. Although this approach ensures a level of consistency and interoperability between the

*Corresponding Author

¹<http://amsterdammuseum.nl>

²<http://www.adlibsoft.com/>

³<http://collectie.ahm.nl>

⁴<http://amdata.adlibsoft.com/wwwopac.ashx>

⁵<http://blog.amsterdammuseum.nl/?p=5245>
(Dutch)

⁶<https://wiki.dnb.de/display/LDS/>

⁷<http://bnb.data.bl.uk>

⁸<http://www.europeana.eu>

⁹<http://data.europeana.eu>

datasets from different institutions it creates a disconnect between the cultural heritage institute original metadata model and the Linked Data version.

Amsterdam Museum Linked Data set was created with a number of design principles in mind. Specifically, we have created a five-star Linked Data set for a small museum that 1) retains the complexity of the original data while 2) achieves interoperability through a mapping to an interoperability layer - in this case the EDM. The Amsterdam Museum Linked Data set implements best practices that -, the together with its methodology and tools- Europeana is keen on adopting for its future workflow.

2. Metadata Conversion and Modeling

2.1. Conversion

We here describe briefly the process used to convert the original data to Linked Data. The methodology and tools that focus on a high level of interactivity and transparency of the process are described in more detail in [1].

The Amsterdam Museum data consists of three parts: 1) an object metadata set consisting of metadata records for the 73.447 objects; 2) a thesaurus consisting of 28.000 concepts used in the metadata records and 3) a person authority file consisting of 66,966 persons related to the objects or the metadata. The metadata, thesaurus and vocabulary were all harvested through an OAI-PMH interface. The resulting XML was first converted to crude RDF and subsequently restructured using interactive rewriting rules. Resources (Objects, concepts, persons, ...) were assigned URIs. For the URI basename, we used <http://purl.org/collections/nl/am/>, here shortened to `am:`. Resource URIs are completed by a unique Amsterdam Museum identifier (the 'preref') with suffix 'proxy-', 'aggregation-', 't-' or 'p-' for proxies, aggregations, concepts and persons respectively (eg. `am:proxy-22476`). For predicate URIs, we use the original XML element names (eg. `am:maker`). We used purl.org URIs since for this conversion we were not in the position to use the Amsterdam Museum namespace for our Linked Data server.

Also in this step, implicit links (for example between between objects and thesaurus concepts) are made explicit as RDF relations. Literal values that represent references to resources are replaced by those resources.

Language information of the literals is also added to the data in the form of language-typed literals.

2.2. Mapping to the Europeana Data Model

To make the Amsterdam Museum Linked Data interoperable with the EDM, two steps are taken. First of all, the museum objects are represented as *proxy-aggregation* pairs. EDM supports multiple providers describing the same object and allows for enrichment of the museum data, while clearly showing the provenance of all the data that links to the digital object. This is achieved by incorporating the proxy-aggregation pattern from the Object Re-use and Exchange (ORE) model¹⁰. For our purpose, this means that an Amsterdam Museum metadata record gives rise to both a proxy resource as well as an aggregation resource. The RDF triples that make up the object metadata (creator, dimensions etc.) have the proxy as their source while the triples that are used for provenance (data provider, rights etc.) as well as digital representation (e.g. thumbnails) have the aggregation resource as their source. An example of this is shown in Figure 2.

The second step is to map the Amsterdam Museum specific classes and properties to those of the EDM using `rdfs:subPropertyOf` or `rdfs:subClassOf` relations. The EDM uses classes and properties from Dublin Core for object descriptions and SKOS for thesaurus descriptions, along with a number of EDM-specific classes and properties. For person metadata, we use RDA Group 2 metadata standard¹¹. These properties include given and family names, birth and death dates etc. As a side note, informed by this conversion, this metadata set is currently considered as the EDM standard for biographical information.

Through these mappings, interoperability of the museum-specific metadata with the EDM is achieved. An example is shown in Figure 1.

3. Description of the Linked Data set

We now describe in more detail the resulting Linked Data set, for each of its three parts. Table 1 lists some statistics. For illustration, Figure 2 shows a small part of the RDF graph for a museum object, including internal links (relations to Amsterdam resources).

¹⁰<http://www.openarchives.org/ore/>

¹¹<http://rdvocab.info/ElementsGr2>

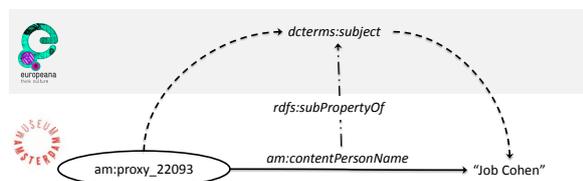


Fig. 1. RDF fragment showing how metadata mapping ensures interoperability. The bottom part of the figure shows an example triple relating an object to the name of a depicted person. Dublin Core (the metadata standard used in Europeana for object descriptions) only has a single notion of the subject of a work. By mapping the specific properties to the more general property using the `rdfs:subPropertyOf` in the metadata schema, an application capable of RDFS reasoning can infer that the object has “Job Cohen” as its subject. We therefore achieve interoperability without discarding the complexity of the original data.

Data part	Resources	Predicates used	RDF triples
Object metadata	73,447 (proxies)	100	5,700,371
Thesaurus	28,000 (concepts)	13	601,819
Person auth. list	66,966 (persons)	21	301,143
Total	168,413	134	6,603,333

Table 1

Some statistics for the three parts of the Linked Data set

3.1. Object Metadata

The object metadata consist of 73,447 proxies-aggregation pairs. The 100 different predicates include creator, dimensions, locations, related exhibitions etc. For complex relations, 566,239 blank nodes were retained. In total the object metadata consists of 5,700,371 RDF triples of which many have a thesaurus concept or person resource as object.

The RDFS mapping file relates the 100 Amsterdam Museum properties to the EDM properties through the `rdfs:subPropertyOf` construct. 90 properties are defined as subproperties of Dublin Core properties, seven properties are mapped to EDM-specific properties (`edm:hasMet`, `edm:happenedAt`, etc.) and three properties are defined as subproperties of `rdfs:label`. Two Amsterdam Museum classes `am:Exhibition` and `am:Locat` were defined as `rdfs:subClassOf` of the EDM class `edm:Event`.

3.2. Thesaurus

The thesaurus consists of 28,000 concepts represented in SKOS. These include geographical terms, motifs, events etc. Most term-based thesauri, including the AM thesaurus, have a more or less uniform structure (ISO 25964) making the representation as SKOS fairly straightforward. For this reason, for most

properties SKOS is directly used. There are 13 predicates including `skos:broader` (7,487 triples) and `skos:narrower` (8,486 triples) that establish hierarchies. In total the thesaurus consists of 601,819 RDF triples (no blank nodes).

3.3. Person authority file

The person authority file consists of 66,966 instances of `am:Person` (a subclass of `edm:Agent`). The persons in this data set are creators, past or present owners of art objects, annotators, depicted persons etc. In this case, the original 21 distinct Amsterdam Museum predicates were used. These predicates include birth and death dates, nationality, alternative name spellings etc. These properties are mapped to RDA Group 2 elements using 20 `rdfs:subPropertyOf` relations were defined. The `am:Person` class was also mapped as a `rdfs:subClassOf` `edm:Agent`. In total there are 301,143 RDF triples in Person data set.

3.4. Internal Links

There are many links between resources in the dataset. There are 558,161 links between the 73,447 proxies to thesaurus concepts, producing an average of 7.6 links to thesaurus concepts per cultural heritage object. There are also 80,432 links between proxies and persons. In addition to this, there are 243,532 links between proxies and other proxies.

3.5. External Links

Links to external data sources were established manually using the Amalgame alignment platform [3]. Again, the alignment process is described in more detail in [1].

The thesaurus was linked partly to the Dutch AATNed¹² thesaurus (2498 links) and partly to GeoNames¹³ (143 links, covering about 75% of the geographic part of the Amsterdam Museum thesaurus). The person authority file was linked to Getty Union List of Artist Names (ULAN)¹⁴, resulting in 1426 links.¹⁵ There are also 34 links to DBpedia.

¹²<http://www.aat-ned.nl>

¹³<http://www.Geonames.org>

¹⁴<http://www.getty.edu/research/tools/vocabularies/ulan>

¹⁵Although ULAN itself is not in the Linked Data cloud, it is incorporated in VIAF, which is in the Linked Data cloud.

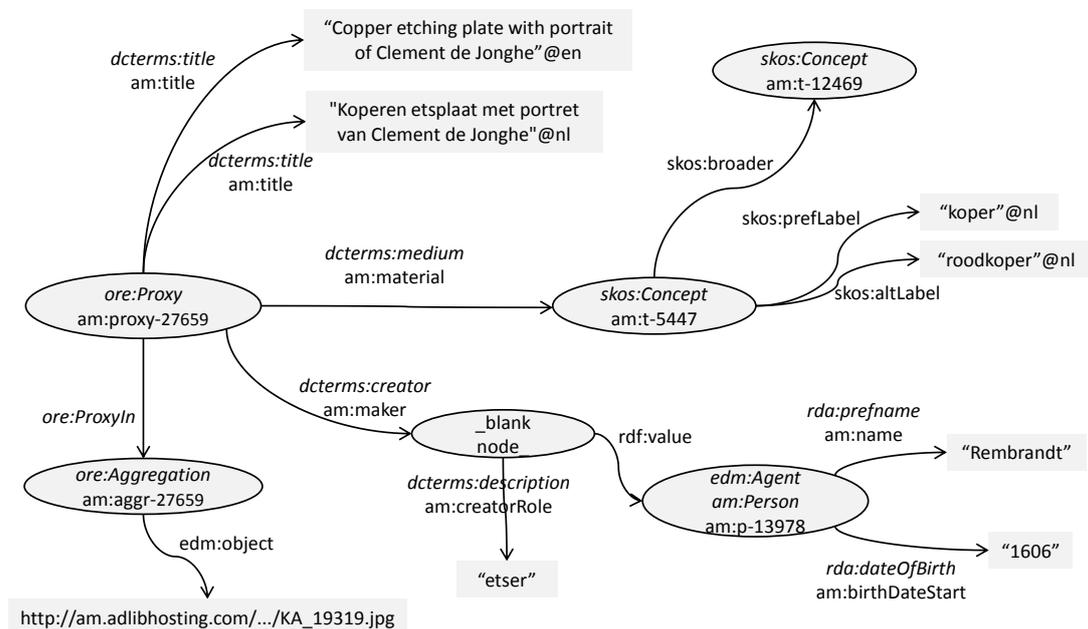


Fig. 2. A small part of the RDF graph surrounding the resource `am:proxy-27659`. Predicates and classes are listed with their super-properties and -classes in italics. On the left side, the relation to the Aggregation (with the attached thumbnail) can be seen. In the top right, the relation to the thesaurus concept is shown. In the bottom right, a complex creator relation (the relation has a type as well as a value) requires a blank node. The `rdf:value` of the blank node is a resource in the Person list. The proxy has two title triples, for the English and Dutch titles respectively, with language-typed literals as objects.

In total 3753 links to external datasources are included. Although this is only a fraction of the total number of concepts, the usefulness of these mappings is much greater as they represent the part of the concepts with which the most metadata is annotated. In total, 70.742 out of the 73.447 (96%) objects are annotated with one or more concepts or persons that have been linked, with an average of 4.3 linked concepts per object. Nevertheless, we are still aiming to enrich the data with more links.

4. Availability

The Amsterdam museum data, consisting of the converted datasets, the schema mapping files and the high-quality mapping files are served as Linked Open Data on the Europeana Semantic Layer (ESL). The ESL is a running instance of the ClioPatria semantic server [4] that houses other datasets that have been mapped to EDM. It can be accessed at <http://semanticweb.cs.vu.nl/europeana>. Amsterdam Museum PURL URIs (for example <http://purl.org/collections/nl/am/proxy-63432>) are redirected to

this server. Based on the response header in the HTTP request, either HTML, RDF/XML or RDF/Turtle is served. The SPARQL endpoint for the ESL is found at <http://semanticweb.cs.vu.nl/europeana/sparql/>¹⁶. In addition to this, the Amsterdam Museum dataset can be separately accessed through a GIT repository at <http://eculture.cs.vu.nl/git/public/?p=econnect/metadata/AHM.git>. More information, including example URIs and SPARQL queries can be found at <http://semanticweb.cs.vu.nl/lod/am>.

5. Discussion

The Amsterdam Museum Linked Data set is a significant data source for Amsterdam history and culture. In previous application development competitions such as Apps for Amsterdam¹⁷, the Amsterdam Museum dataset has been used extensively for a num-

¹⁶An interactive SPARQL query page is available at <http://semanticweb.cs.vu.nl/europeana/user/query>

¹⁷<http://www.appsforamsterdam.nl/>

ber of mashup applications and we expect that the Linked Data version will be an equally central data set for the Linked Culture Data web. We are currently developing a mobile cultural tour guide for which the Amsterdam Museum Linked Data set will be a central datasource.

The Amsterdam Museum Data Set was developed in the context of the EuropeanaConnect¹⁸ research project. It serves as a prototype Linked Data set in which the original richness of the data is maintained, while still being interoperable through its mapping to the EDM. A limitation is that this interoperability does require some RDFS reasoning by applications wishing to access the data at the interoperability level.

Where the current Linked Data pilot of Europeana (data.europeana.eu) focusses on producing a Linked Data set based on the already-ingested metadata consisting of a minimal set of Dublin Core properties. Current research within Europeana is gearing towards harvesting and producing more complex data sets, in the manner of the Amsterdam Museum one. When more of these data sets will appear, more links to Amsterdam Museum data are likely.

In the current workflow, the Linked Data is not updated automatically once a change has been made in the original collection management system. The OAI-

PMH harvesting, conversion and data uploading steps would have to be run again. Since the conversion rules are retained, this can be done periodically in an automated fashion. The conversion tool ensures that the same resources receive the same URIs.

References

- [1] V. de Boer, J. Wielemaker, J. van Gent, M. Hildebrand, A. Isaac, J. R. van Ossenbruggen, and G. Schreiber. Supporting Linked Data Production For Cultural Heritage Institutes: The Amsterdam Museum Case Study. In *Proceedings of European Semantic Web Conference (ESWC)*, 2012.
- [2] Antoine Isaac. Europeana data model primer. <http://pro.europeana.eu/edm-documentation>, 2010.
- [3] Jacco van Ossenbruggen, Michiel Hildebrand, and Viktor de Boer. Interactive vocabulary alignment. In Stefan Gradmann, Francesca Borri, Carlo Meghini, and Heiko Schuldt, editors, *TPDL*, volume 6966 of *Lecture Notes in Computer Science*, pages 296–307. Springer, 2011.
- [4] Jan Wielemaker, Michiel Hildebrand, Jacco Ossenbruggen, and Guus Schreiber. Thesaurus-based search in large heterogeneous collections. In *Proceedings of the 7th International Conference on The Semantic Web, ISWC '08*, pages 695–708, Berlin, Heidelberg, 2008. Springer-Verlag.

¹⁸<http://www.europeanaconnect.eu>