Reason-able View of Linked Data for Cultural Heritage

Mariana Damova[†], Dana Dannélls[§]

[†]Ontotext, Tsarigradsko Chausse 135, Sofia 1784, Bulgaria
[§]University of Gothenburg, Lennart Torstenssonsgatan 8, 405 30 Gothenburg, Sweden
mariana.damova@ontotext.com, dana.dannells@svenska.gu.se

Abstract. This paper presents a novel approach that relies on the innovative idea of Reason-able View of the Web of linked data applied to the domain of cultural heritage. We describe an application of data integration based on Semantic Web technologies and the methods necessary to create an integrated semantic knowledge base composed of real museum data that are interlinked with data from the Linked Open Data (LOD) cloud. Thus, creating an infrastructure to allow for easy extension of the domain specific data, and convenient querying of multiple datasets. Our approach is based on a model of schema level and an instance level alignment. The models use several ontologies, e.g. PROTON and CIDOC-CRM, showing their integration *by using* real data from the Gothenburg City Museum.

Keywords: Linked Open Data, Reason-able View, cultural heritage, museum, ontology, data integration, Semantic Web

1 Introduction

Being able to obtain useful information from Linked Open Data (LOD) [12], i.e. combining knowledge and facts from different datasets is the ultimate goal of the Semantic Web. Although clear, the vision of LOD and the Semantic Web is still looking for convincing real life use cases demonstrating the benefits of these technologies. MacManus in [13] defines one exemplar test for the Semantic Web. He formulates a conceptual query about cities around the world which have "Modigliani artwork", and states that the vision of the Semantic Web will be realized when an engine will return an answer to it. Actually, the answer to this question can be found in the LOD; where different facts about the artist, his artwork and the museums or galleries that host them are to be found in different datasets. To our knowledge FactForge [4], a public service provided by Ontotext, is the only engine capable of passing this test (cf. Fig. 1).[1] FactForge is based on the method of *Reason-able Views* of the web of data [9], [10].

[1] The SPARQL query to obtain this information can be run at http://factforge.net/sparql.

SPARQL Query

Results for PREFIX fb: <http://rdf.fr... (12) View as Exhibit Download in JSON | SPARQL Results in XML | SPARQL Results in JSON

painting_l	owner_l	city_fb_con	city_db_loc	city_db_cit
Head@en	Museum of Modern Art	Manhattan		
Head@en	Museum of Modern Art	New York City		
Anna Zborowska@en	Museum of Modern Art	Manhattan		
Anna Zborowska@en	Museum of Modern Art	New York City		
Portrait of Diego Rivera@en	The São Paulo Museum of Art@en		Sao Paulo@en	
Portrait of Diego Rivera@en	The São Paulo Museum of Art@en		São Paulo	
Woman with a Necklace@en	School of the Art Institute of Chicago@en			Chicago
Portrait of a Woman@en	School of the Art Institute of Chicago@en			Chicago
Reclining Nude@en	Museum of Modern Art	Manhattan		
Reclining Nude@en	Museum of Modern Art	New York City		
Madam Pompadour@en	School of the Art Institute of Chicago@en			Chicago
Jeanne Hébuterne@en	Barnes Foundation@en	Philadelphia		

Figure 1. Results of the Modigliani test.

The "Modigliani artwork" example gives evidence for the potential of the cultural heritage domain to become a useful use case for the application of the semantic technologies. Our work is a step in this direction showing a Reason-able View of the web of data integrating museum and LOD cloud data. In this paper we present a Reason-able View of the web of data, using real museum data that are integrated with data from the LOD cloud.

2 Linked Open Data - the Vision

The notion of "linked data" is defined by Tim Berners-Lee, [1] as RDF [14] graphs, published on the WWW and explorable across servers in a manner similar to the way the HTML web is navigated. Linked Open Data (LOD) is a W3C SWEO community project aiming to extend the Web by publishing open datasets as RDF and by creating RDF links between data items from different data sources, cf. Fig. 2.

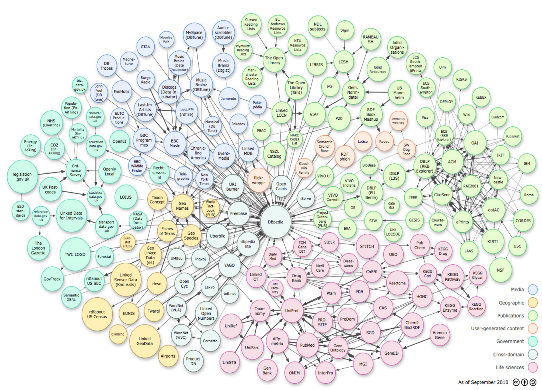

Figure 2 The LOD Cloud.

3 Linked Open Data Management

Using linked data for data management is considered to have great potential in view of the transformation of the web of data into a giant global graph [3]. Kiryakov et al., [10] present a *Reason-able View (RAV)* approach for reasoning with and managing linked data. RAV is an assembly of independent datasets, which can be used as a single body of knowledge with respect to reasoning and query evaluation. It aims at lowering the cost and the risks of using specific linked datasets for specific purposes. The linkage between the data is made at the schema-level by mapping ontologies [3], [8], and at the instance level with the predicate `owl:sameAs`, i.e. the common method of connecting data in the LOD cloud. *Reason-able Views* are accessible via a SPARQL [15] end-point and keywords. Because each Reason-able View is a compound dataset, i.e. it consists of several datasets, and one can formulate queries, combining predicates from different datasets and ontologies in a single SPARQL query. The results from such queries return instances which also come from different datasets in the Reason-able View.

4 Museum Reason-able View

The datasets in each Reason-able View depend on the underlying purpose of use of the compound dataset. In our case, the Museum Reason-able View has to be constructed in a way to provide adequate content for the two following requirements:

- the ability to handle generic knowledge, such as people, institutions, and locations
- the ability to handle specific subject domains, such as the cultural heritage and museums

The Museum Reason-able View, presented in this paper, comprises a heterogeneous dataset reflecting a combination of generic knowledge, and domain specific knowledge. It includes the following datasets from the LOD cloud:

- **DBpedia**[2] - the RDF-ized version of Wikipedia, describing more than 3.5 million things and covers 97 languages.
- **Geonames**[3] - a geographic database that covers 6 million of the most significant geographical features on Earth.
- **PROTON**[4] - an upper-level ontology, 542 entity classes and 183 properties.

These datasets cover the generic knowledge of the Museum Reason-able View.

The next sections introduce the Museum specific knowledge integrated into the Museum Reason-able View.

---

[2] DBPedia, structured information from Wikipedia: http://dbpedia.org.

[3] Geonames, a geographical database: http://www.geonames.org.

[4] PROTON, a lightweight upper-level ontology: http://proton.semanticweb.org/.

5 Museum Data Models

The CIDOC-CRM is an object oriented ontology developed by the International Council of Museum's Committee for Documentation (ICOM-CIDOC)[5], with overall scope of curated knowledge of museums. The model provides the level of details and precision necessary for museum professionals to perform their work well. The CIDOC-CRM ontology consists of about 90 classes and 148 properties. It represents an upper-level ontology view for cultural and natural history. Its higher level concepts are general concepts, e.g. `Entity`, `Temporal Entity`, `Time Span`, `Place`, `Dimension`, `and Persistent Item`. Physical items and non-material products produced by humans are described as `Man-made-thing`, and `Conceptual Object`. The concept `Event` of CIDOC-CRM covers through its sub concepts the entire lifecycle of an artifact, e.g. `Production`, `Creation`, `Dissolution`, `Acquisition`, `Curation`, etc. Some of these concepts have more than one immediate superclass.

The integration of CIDOC-CRM into the Museum Reason-able View takes place at the schema level by providing mappings between the CIDOC-CRM concepts and PROTON concepts, cf. Fig. 4. The CIDOC-CRM concepts are linked to PROTON concepts with the built-in property `owl:equivalentClass.` Six classes from CIDOC-CRM and PROTON are being interlinked in this way.

K-samsök [11], the Swedish Open Cultural Heritage (SOCH), is a Web service for applications to retrieve data from cultural heritage institutions or associations with cultural heritage information. The idea behind K-samsök is to harvest any data format and structure that is used in the museum sector in Sweden and map it into K-samsök's categorization structure available in an RDF compatible form. It includes features which are divided in the following categories:

(a) Identification of the item in the collection
(b) Internet address, and thumbnail address
(c) Description of the item
(d) Description of the presentation of the item, including a thumbnail
(e) Geographic location coordinates
(f) Museum information about the item
(g) Context, when was it created, to which style it belongs, etc.
(h) Item specification, e.g. size, and type of the item – painting, sculpture and the like

Fig. 3 presents a painting item from The History Museum in Sweden described according to these categories available at the following URL:

`http://mis.historiska.se/mis/sok/fid.asp?fid=96596&g=1`

[5] CIDOC CRM webpage: http://www.cidoc-crm.org/.

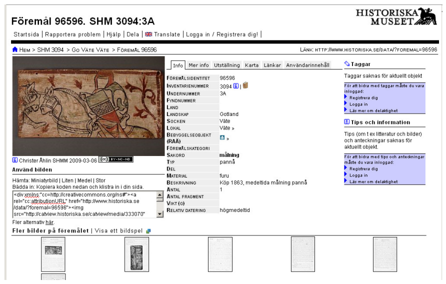

Figure 3 Website of a painting item from the History Museum in Sweden.

The CIDOC-CRM schema is not enough to cover all the information that K-samsök tends to capture. In order to provide the necessary infrastructure to load the complete information about a museum item, it is required to integrate the schema of K-samsök into the Museum Reason-able View. This is possible by defining a new intermediary layer described in a specific ontology, which we will call the Museum Artifacts Ontology (MAO).[6] The MAO ontology was developed for mapping between museum data and the K-samsök schema. The ontology includes concepts reflecting the K-samsök schema to allow integrating the data from the Swedish museums. It has about 10 concepts and about 20 new properties.

It is important to note, that this Museum Artifacts Ontology can be further specified with descriptions of additional concepts covering a specific type of museum artifacts, like for example paintings.

6 The Gothenburg Museum data

The Gothenburg museum [5] preserves 8900 museum objects described in its database. These objects correspond to two museum collections (GSM and GIM) and are placed in two tables of the museum database. 39 concept fields display each museum object, including its identification, its type - a painting, a sculpture, etc.-, its material, its measurements, its location, etc. All concept fields are described in Swedish.

The Gothenburg City Museum database structure follows the structure of the CIDOC-CRM, and the part of its data described above is used as experimental data for the Museum Reason-able View. The data is mapped to concepts from PROTON and MAO in the cases when the concepts available in the data are not available in CIDOC-CRM. Fig. 4 shows the architecture of the integration of the Gothenburg City Museum data into the Museum Reason-able View by representing and linking them with elements from different schemata, e.g. PROTON, CIDOC-CRM and MAO.

Additionally, the linkage with external to the Gothenburg City Museum data, e.g. DBpedia, is provided by connecting the MAO concepts to DBpedia instances, or by

6 It is just a coincidence that this ontology has the same as the Finish MAO [6], which also describes museum artifacts for the Finish museums.

connecting the Gothenburg museum data with the corresponding DBpedia instances using the predicate `owl:sameAs`.

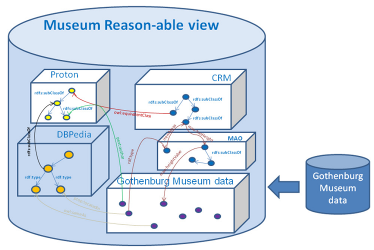

Figure 4 Dataset interconnectedness in the Museum Reason-able View.

The process of the Gothenburg City Museum data integration into the Museum Reason-able View consists in transforming the information from the museum database into RDF triples based on the described ontologies. Each museum item is given an unique URI, and the concept fields from the database are interpreted as describing concepts or properties from one of the three ontologies, e.g. PROTON, CIDOC-CRM or MAO. The objects of the triples are derived from the columns of the database.

The triple generation goes through a process of localization, e.g. using English words for the naming of the properties and URIs in the Museum Reason-able View.

Loading the Gothenburg City Museum data into the Museum Reason-able View enables queries of the following nature:

- Museum artefacts preserved in the museum since 2005
- Paintings from the GSM collection
- Inventory numbers of the paintings from the GSM collection
- Location of the objects created by Anders Hafrin
- Paintings with length less than 1 meter
- etc.

7 Museum Reason-able View Environment

The Museum Reason-able View environment is built as an instance of BigOWLIM triple store. It provides the knowledge to query Gothenburg City Museum data in a structured way. It contains: DBPedia 3.6, Geonames, PROTON, CIDOC-CRM and MAO ontologies, and their mappings, and the triplified Gothenburg City Museum data. BigOWLIM performs full materialization during loading. It was expected that the available retrievable statements after loading will exceed the loaded explicit statements by about 20%. The loading statistics confirmed this expectation, e.g. the number of the loaded explicit statements was 257,774,678 triples, whereas the overall number of triples available for querying was 16% more, e.g. 305,313,536.

8 Related Work

Museum Data Integration with semantic technologies as proposed in this paper is intended to enable efficient sharing of museum and cultural heritage information. Initiatives about developing such sharing museum data infrastructures have increased in the recent years. Only few of them rely on semantic technologies. Similar project has been carried out for the Amsterdam Museum, developed by VUA.[7] This project aims at producing Linked Data within the Europeana[8] data model. To our knowledge ours is the first attempt of using CIDOC-CRM to produce museum linked data with connections to external sources from the LOD cloud like DBpedia and Geonames. Schema-level alignment is a new method of achieving interoperability in LOD [3], [8]. This method has not been applied on data in the cultural heritage domain, which we propose in this paper.

9 Conclusion

We presented the methods of using a knowledge representation infrastructure to build a knowledge base in the cultural heritage domain according to the described above innovative methods and models. The Museum Reason-able View provides an easy path to extension of the knowledge base with data from other Swedish museums or generally museum data, and allows to query and obtain results not only about artifacts belonging to different museum collections but also general knowledge about them from DBpedia and Geonames.

Our future work includes detailed experiments with the Museum Reason-able View regarding querying and navigation, extensions of the data models to cover detailed museum artifacts descriptions, like paintings, and using the interlinked ontologies as an interface for access to and presentation of the structured museum data in natural language.

Acknowledgments. This work is supported by MOLTO European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement FP7-ICT-247914.

References

1. Berners-Lee, T.: Design Issues: Linked Data. http://www.w3.org/DesignIssues/LinkedData.html
2. Bishop, B., Kiryakov, A., Ognyanoff, D., Peikov, I., Tashev, Z., Velkov, R. : OWLIM: A family of scalable semantic repositories. Semantic Web Journal, Special Issue: Real-World Applications of OWL. http://www.semantic-

[7] VUA: http://www.vu.nl/en/.
[8] Europeana portal: http://www.europeana.eu/portal/.

web-journal.net/content/new-submission-owlim-family-scalable-semantic-repositories

3. Damova, M., Kiryakov, A. Grinberg, M., Bergman, M., Giasson, F., Simov, K. Creation and Integration of Reference Ontologies for Efficient LOD Management. In: Semi-Automatic Ontology Development: Processes and Resources, IGI Global, Hershey PA, USA, 2011. (to appear)

4. FactForge, a Reason-able View to the Web of Data. http://factforge.net, http://www.ontotext.com/factforge.

5. Gothenburg City Museum. http://www.stadsmuseum.goteborg.se/prod/kultur/stadsmuseum/dalis2.nsf/0/3955ae522e06a42bc12573950046ec10?OpenDocument.

6. Hyvyonen, E., Maekelae, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S. (2008) Museum Finland. Finnihs Museum on the Semantic Web. www.websemanticsjournal.org/index.php/ps/article/download/74/72.

7. International Council of Museums (ICOM). http://icom.museum/.

8. Jain, P., Yeh, P. Z., Verma, K., Vasquez, R. G., Damova, M., Hitzler, P., and Sheth, A. P. (2011). Contextual Ontology Alignment of LOD with an Upper Ontology: A Case Study with Proton. In G. Antoniou (Ed.), *Proceedings of 8th Extended Semantic Web Conference.* Heraklion, Crete.

9. Kiryakov, A., and Momtchev, V. *Two Reason-able Views to the Web of Linked Data.* Presentation at the Semantic Technology Conference 2009, San Jose. http://www.slideshare.net/ontotext/two-reasonable-views-to-the-web-of-linked-data.

10. Kiryakov, A., Ognyanoff, D., Velkov, R., Tashev, Z., and Peikov, I. (2009) LDSR: Materialized Reason-able View to the Web of Linked Data. In Rinke Hoekstra and Peter F. Patel-Schneider (Eds) *Proceedings of OWLED 2009*: Chantilly, USA.

11. K-samsök http://www.ksamsok.se/om-k-samsok/anslutna-institutioner/.

12. Linking Open Data, W3C Semantic Web Education and outreach community project, http://linkeddata.org/.

13. MacManus, R. The Modigliani Test: The Semantic Web's Tipping Point, Retrieved April, 2010, http://www.readwriteweb.com/archives/the_modigliani_test_semantic_web_tipping_point.php.

14. RDF Primer. http://www.w3.org/TR/rdf-syntax/.

15. SPARQL. The Query Language for RDF. http://www.w3.org/TR/rdf-sparql-query/.