

Accessing Cultural Heritage using the Web of Data

Jacek Jankowski, Yolanda Cobos, Michael Hausenblas and Stefan Decker

Digital Enterprise Research Institute, NUI Galway, Ireland

Abstract

Cultural Heritage (CH) is a vast domain, where sharing information is challenging. As a result, the global CH is distributed and heterogeneous. In this paper, we present the concept of the Web of Data as an approach to integrating CH collections. We introduce the CHoWDer (Cultural Heritage on the Web of Data) model, the simple model designed to illustrate how easily CH institutions can participate in the Linked Data initiative.

Categories and Subject Descriptors (according to ACM CCS): J.5 [Computer Application]: Arts and Humanities—H.3.7 [Information Systems]: Information Storage and Retrieval—Digital Libraries

1. Introduction

The number of audiovisual assets owned by Cultural Heritage institutions has increased along with the evolution of multimedia technologies. Putting these collections on the Web, providing search and access to these collections to broad classes of users, seems to be a natural solution for their dissemination. However, Cultural Heritage (CH) is a vast domain, where sharing information is challenging - the sources contain a variety of types of structured and unstructured resources, combining text, images, videos and 3D objects. It is very costly to be compliant with standards for describing such diverse content. As a result, the global CH is distributed among isolated islands.

This leads to a problem with searching for information. Where should the CH users search, what are their information needs? The key findings of the user study on the information seeking behavior of CH experts and the sources they use to carry out search tasks include [AvOHvN08]:

- the majority of search tasks involves relatively complex information gathering (in contrast to the simple fact finding oriented support provided by current tools);
- many search tasks require experts to use and combine results from multiple sources.

The questions arise: Is it possible to improve and accelerate the data exploration process? How to integrate CH collections? In this article we try to answer these questions. Our work was inspired by the two great ideas: (1) the work of Doerr et al. and their *Dream of Global Knowledge Network*

(Section 2) and (2) the Tim Berners-Lee's Linked Data (LD) initiative (Section 3). The former offers a breakthrough to problems of semantic interoperability in the CH domain. The later refers to a style of publishing and interlinking structured data on the Web. The aim of this paper is to bring the two ideas together and to discuss how to share CH content on the Web along with implementing the linked data principles. The contribution of this paper is the CHoWDer (Cultural Heritage on the Web of Data) model (Section 4), a simple approach to interlinking collections of different CH institutions.

2. Global Knowledge Network

Martin Doerr and his colleagues believe that it is time to revive the dream of a global network of knowledge (after semantic networks, the Web and the Semantic Web) [DHL03, DI08], a knowledge base, that will integrate complementary CH information sources [DST04].

In [DI08] they present the CIDOC Conceptual Reference Model (CRM) as a nearly generic semantic model for summarizing, structuring, and combining existing data. It aims at building a distributed platform for cultural resources, which consist of databases and repositories of different CH institutions and, at the same time, supporting queries as in a whole.

"The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended

to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives." [CIDOC CRM Home Page]

Linked Data, comparing to the vision of a global knowledge network (which focus on the ontological level and inferencing), is mainly about publishing structured data in RDF using URIs. It is designed to significantly lower the barrier for authoring semantic content on the Web.

3. Linked Data

With linked data [BLa] the idea of a hyperlinks is re-introduced in the RDF-based Knowledge Representation world. We will discuss the underlying principles, available datasets following the principles and point out the benefits in the following.

3.1. Linked Data Principles

In 2006, the inventor of the Web, Tim Berners-Lee, outlined the four principles of linked data [BLa]:

- Use URIs as names for resources;
- Use dereferenceable URIs (that is, HTTP URIs) so that people can look up those names;
- When someone looks up a URI, provide useful information (representations are assumed to be RDF);
- Links to resources in other datasets should be included in order to enable the discovery of more data.

Adhering to these four linked data principles makes an opportunity to make data interconnected. We note that the linked data principles roughly constitute a simplified (in the sense of read-only) RESTful architecture.

3.2. Linking Open Data

From 2007 on, the W3C-initiated community project "Linking Open Data" (LOD) aimed at applying the linked data principles to publicly available datasets, such as Wikipedia, DBLP, etc.—currently, over 100 datasets with billions of triples and millions of interlinks have already been made available by numerous institutions and projects (Figure 1), incl. the BBC, life science groups, as well as individuals.

In the next section we will highlight the benefits of linking data and discuss how to exploit LOD for certain domains [Hau09].

3.3. Benefits of Linking Data

In [HvOT*09], Hardman et. al. discuss potential uses of linked data to support users' information needs and give examples of using linked data to support user information seeking tasks:

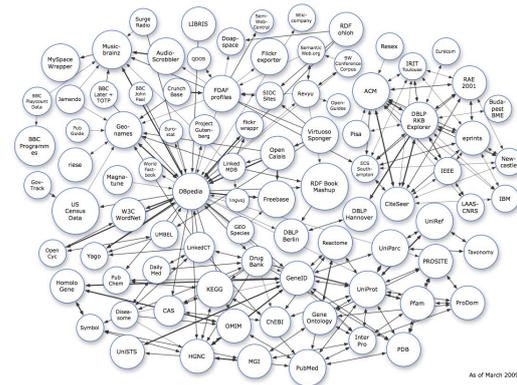


Figure 1: LD Datasets (from <http://linkeddata.org>)

- Exploring cultural heritage repositories: the underlying linked data can improve search results and their presentation; more results can be retrieved for a single term query because of the underlying relations linking terms through thesauri incorporated within the infrastructure.
- Developing support for identifying annotation terms: the role of linked data there was to connect terms occurring in different thesauri, saving time by using a single search.

Linked Data has already been applied to multimedia content, for example, "Catch Me If You Can" (CaMi-Catzee) [HH08], a multimedia interlinking concept demonstrator allows people to semantically annotate pictures on Flickr. Such an annotation interlinks the picture to a resource containing more information about the depicted person (e.g., a homepage, where the person's FOAF profile is stored).



Figure 2: DBpedia Mobile's map view [BB08].

DBpedia Mobile is a location-centric DBpedia client application for mobile devices consisting of a map view and a Linked Data browser [BB08]. The DBpedia project extracts structured information from Wikipedia and publishes this information as Linked Data on the Web. The DBpedia dataset contains information about 2.18 million things, including almost 300,000 geographic locations. It is interlinked with various other location-related datasets. Based on the current GPS position of a mobile device, DBpedia Mobile can render a map indicating nearby locations from the DBpedia dataset (see Figure 2).

A good place to start to appreciate the benefits of linked data is Tim Berners-Lee's TED talk [BLb].

4. CHoWDer Model

In order to become a part of the Web of Data, CH institutions only have to design their applications according to the Linked Data principles (see Section 3.1). Metadata descriptions have to be interoperable in order to reference and integrate parts of the described CH resources.

In this section we describe the CHoWDer (Cultural Heritage on the Web of Data) model, a simple approach to interlinking collections of different CH institutions (see Figure 3). Its purpose is to illustrate how easily any institution can participate in the Linked Data initiative.

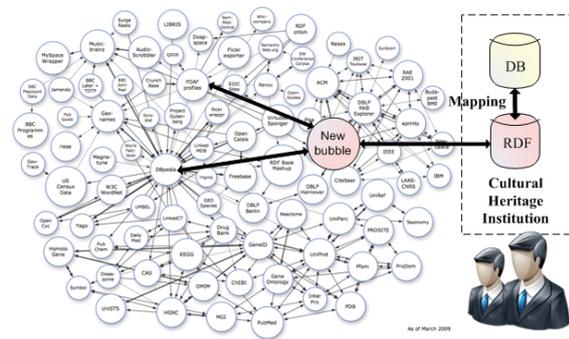


Figure 3: The CHoWDer model.

Typically, any CH institution store information about its collections in some kind of repository (DB on the Figure 3) - it is usually an SQL database. Let us assume that the database stores the information about Epitaphios GE34604 (we use data example for the CRM from the museum Benaki, see CIDOC Home Page; small part of this data is presented on Figure 4).

```
<CRMset>
  <CRM_Entity>Epitaphios GE34604
    <in_class>E23: Information Carrier</in_class>
    <is_identified_by>GE 34604
      <in_class>E42: Object Identifier</in_class>
    </is_identified_by>
    <is_documented_in>
      http://cidoc.ics.forth.gr/graphics/epit1.jpg
    <in_class>E31: Document</in_class>
    </is_documented_in>
    <current_owner>Museum Benaki !!!
      <in_class>E40: Legal Body</in_class>
    </current_owner>
  </CRM_Entity>
</CRMset>
```

Figure 4: CRM description of Epitaphios GE34604.

Let us assume that this data is mapped into RDF (Resource Description Framework) and made available online (a critical review of various methods and tools that can be

used for interlinking resources in the Web of Data can be found in [HTRB09]). Adding a triple:

```
<http://CHInstitution.org/MuseumBenaki>
  owl:sameAs
  <http://dbpedia.org/resource/Benaki_Museum>
```

does not interfere with data. However, as our RDF graph share the URI http://dbpedia.org/resource/Benaki_Museum with DBpedia graph, they merge together, as shown on Fig 5.

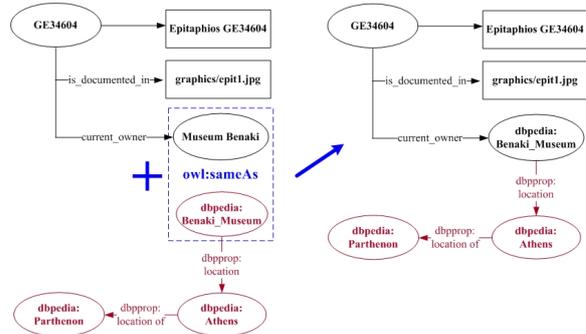


Figure 5: Linking from CIDOC-CRM to DBpedia.

As a result, any user interested in Epitaphios GE34604 will have an access to the RDF graph that contains RDF links to related information, for instance, that the painting is in Museum Benaki, which is located in Athens, etc.

Our model was not designed to answer all questions regarding linking CH to the Web of Data; rather, to explore possibilities of interlinking data. In [BCH] Bizer et al. provide a tutorial on how to publish Linked Data on the Web. In [KLS*07] Kakali et al. present a methodology for mapping DCMI Type vocabulary to CIDOC/CRM, which can be of great use to CH experts, since Dublin Core is widely used on the Web of Data.

We are implementing the CHoWDer model in 3dWiki (see <http://www.3dwiki.org>), the wiki engine developed for Copernicus Encyclopedia [JD09]. The relational database of 3dWiki can store text, images, videos and 3D objects together with metadata. For serving Linked Data views on our database we use D2R Server [BC]. D2R server relies on a declarative mapping between the schemata of the database and the target RDF terms. Based on this mapping, D2R Server serves a Linked Data view on your database and provides a SPARQL endpoint for it (more information can be found on <http://www.3dWiki.org/>).

5. Related Work

Related work to ours can be traced back to hypermedia research [HvOM*99, HBvR93]. We note that the underlying technologies used in the Web of Data setup (such as RDF, RDF-based vocabularies, etc.) usually differ significantly

from what has been assumed to form the basis in these early works. However, there is also a number of recent works, much more aligned to the RDF-based Web of Data, we have identified to be relevant to our approach, discussed in the following.

In [Rob] Roberto describes an idea to open up the data that museums hold about their objects, and to expose the data. He points out, that: "*the work of creating user interfaces does not even have to be done by the museum itself - if the data is made public, interested third parties can interpret the data themselves. This spirit of openness allows museum objects to form part of the "Web of Data" that is now emerging, and holds great potential to engage the public in the debate about the history and future of museum collections.*"

In [SAA*08], Schreiber et. al. describe the MultimediaN E-Culture demonstrator, a Semantic Web application for semantic annotation and search in large virtual collections of cultural-heritage objects, indexed with multiple vocabularies. The system was used to explore how the relations in linked data can be used to improve the presentation of results at the user interface [HvOT*09].

In [REMA09], Rodriguez-Echavarría et al. present how to integrate 3D content with semantic information. They show mechanism for tagging, which links parts of a 3D geometry to CIDOC-CRM URIs in order to store entry points to a metadata repository. Like our 3dWiki, their architecture uses the D2R server.

In [HBM07], Hausenblas et al. describe a multimedia metadata deployment framework based on RDFa, called `ramm.x`. As Google announced support for RDFa only recently, we believe that this is a promising approach to expose—from a Web of Data point of view—legacy multimedia metadata formats, such as MPEG-7 or Exif on the Web of Data.

6. Conclusions and Future Work

The mission of Digital Enterprise Research Institute, the place where we work, is "*Enabling networked knowledge*". We aim to exploit semantics for people, organizations, and systems to collaborate and interoperate on a global scale. The CH is a legacy from the past which we can pass on to current and future generations by linking it to the Web of Data. Without CH the global network of knowledge will be.

In this article, we presented ChoWDer, the simple model designed to illustrate how easily CH institutions can participate in the Linked Data initiative. The future work is to explore possibilities of reusing interlinked data; we believe that there are many unexpected and interesting ways to reveal.

References

[AvOHvN08] AMIN A., VAN OSSENBRUGGEN J., HARDMAN L., VAN NISPEN A.: Understanding cultural heritage experts' information seeking needs. In *JCDL* (2008). 1

- [BB08] BECKER C., BIZER C.: Dbpedia mobile: A location-enabled linked data browser. In *LDOW* (2008). 2
- [BC] BIZER C., CYGANIAK R.: D2r server - publishing relational databases on the semantic web. <http://www4.wiwiwiss.fu-berlin.de/bizer/d2rserver/>. 3
- [BCH] BIZER C., CYGANIAK R., HEATH T.: How to publish linked data on the web. <http://www4.wiwiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>. 3
- [BLa] BERNERS-LEE T.: Linked data, design issues. <http://www.w3.org/DesignIssues/LinkedData.html>. 2
- [BLb] BERNERS-LEE T.: The next web. <http://www.ted.com/>. 3
- [DHL03] DOERR M., HUNTER J., LAGOZE C.: Towards a core ontology for information integration. *Journal of Digital Information* (2003). 1
- [DI08] DOERR M., IORIZZO D.: The dream of a global knowledge network - a new approach. *ACM Journal on Computers and Cultural Heritage* (2008). 1
- [DST04] DOERR M., SCHALLER K., THEODORIDOU M.: Integration of complementary archaeological sources. In *CAA* (2004). 1
- [Hau09] HAUSENBLAS M.: Exploiting Linked Data to Build Web Applications. *IEEE Internet Computing* (2009). 2
- [HBM07] HAUSENBLAS M., BAILER W., MAYER H.: Deploying Multimedia Metadata in Cultural Heritage on the Semantic Web. In *First International Workshop on Cultural Heritage on the Semantic Web* (2007). 4
- [HBvR93] HARDMAN L., BULTERMAN D. C. A., VAN ROSSUM G.: Links in hypermedia: the requirement for context. In *HYPertext* (1993). 3
- [HH08] HAUSENBLAS M., HALB W.: Interlinking multimedia data. In *I-Semantics* (2008). 2
- [HTRB09] HAUSENBLAS M., TRONCY R., RAIMOND Y., BÜRGER T.: Interlinking Multimedia: How to Apply Linked Data Principles to Multimedia Fragments. In *LDOW* (2009). 3
- [HvOM*99] HARDMAN L., VAN OSSENBRUGGEN J., MULLENDER K. S., RUTLEDGE L., BULTERMAN D. C. A.: Do you have the time? composition and linking in time-based hypermedia. In *HYPertext* (1999). 3
- [HvOT*09] HARDMAN L., VAN OSSENBRUGGEN J., TRONCY R., AMIN A., HILDEBRAND M.: Interactive information access on the web of data. In *WebSci: Society On-Line* (2009). 2, 4
- [JD09] JANKOWSKI J., DECKER S.: 2lip: Filling the gap between the current and the three-dimensional web. In *Web3D* (2009). 3
- [KLS*07] KAKALI C., LOURDI I., STASINOPOULOU T., BOUNTOURI L., PAPTAEODOROU C., DOERR M., GERGATSOULIS M.: Integrating dublin core metadata for cultural heritage collections using ontologies. In *DCMI* (2007). 3
- [REMA09] RODRIGUEZ-ECHAVARRIA K., MORRIS D., ARNOLD D.: Web based presentation of semantically tagged 3d content for public sculptures and monuments in the uk. In *Web3D* (2009). 4
- [Rob] ROBERTO F.: Exploring museum collections on-line: The quantitative method. In *Museums and the Web 2008*. 4
- [SAA*08] SCHREIBER A., AMIN A., AROYO L., VAN ASSEM M., DE BOER V., HARDMAN L., HILDEBRAND M., OMELAYENKO B., VAN OSSENBRUGGEN J., TORDAI A., WIELEMAKER J., WIELINGA B.: Semantic annotation and search of cultural-heritage collections: The multimedien demonstrator. *Journal of Web Semantics* (2008). 4