

Linked Open Data for Cultural Heritage: Evolution of an Information Technology

Jeff Edelstein, Lola Galla, Carolyn Li-Madeo, Julia Marden, Alison Rhonemus, Noreen Whysel

Spring 2013

Abstract

This paper surveys the landscape of linked open data projects in cultural heritage, examining the work of groups from around the world. Traditionally, linked open data has been ranked using the five star method proposed by Tim Berners-Lee. We found this ranking to be lacking when evaluating how cultural heritage groups not merely develop linked open datasets, but find ways to use linked data to augment user experience. Building on the five-star method, we developed a six-stage life cycle describing both dataset development and dataset usage. We use this framework to describe and evaluate fifteen linked open data projects in the realm of cultural heritage.

Introduction

Linked Open Data (LOD) projects are happening all around the world, expanding the way that we access cultural heritage. Libraries, museums, and archives are figuring out new ways to export their data in triples, integrate external linked datasets into their collections, and develop new interfaces for users to experience cultural heritage. What is Linked Open Data? Many other papers have defined the concept (see, e.g., Berners-Lee, 2006), so this report will not reiterate these definitions in great detail. We approach linked open data as a philosophy as much as a technological format. We borrow heavily from the Five Stars of Linked Open Data as conceptualized by Tim Berners-Lee:



Figure 1: Is Your Linked Open Data 5-Star? (Berners-Lee, 2006)

Berner's Lee's Linked Open Data systems starts with four rules:

1. Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

(Berners-Lee, 2006)

Unless the system complies with all four rules, it is not considered to be linked data. If any of these rules is not followed, then the data cannot be interlinked or reused, a key idea for linked data. Increasingly, new URI schemes are being created. One of the most widely used URI schemes is DBpedia, which is based on data in Wikipedia. The basic format for serving URIs is RDF/XML, a universal format that allows data exchange between systems. Linked Open Data is linked data that is freely available to access, download, and use. It is distributed on an open license (1-star); as machine-readable, structured data (2-star); in a nonproprietary format (3-star); World Wide Web Consortium (W3C; 4-star); and is linked to other peoples data (5-star). In researching cultural heritage datasets, it became clear to us that these types of institutions are mostly somewhat immature relative to Berners-Lees five-star system. Data may be linked but not open, or open but not thoroughly linked. A few institutional projects, notably those of national library systems such as the Library of Congress, are much closer to Berners-Lees definition of linked open data. We wanted to create a system that addresses the reality of these conceptual systems and their stage of development toward the ideal of a 5-star-compliant linked open data project.

A linked dataset converts a basic catalog of cultural heritage items into RDF triples using a predefined vocabulary. These datasets can then be matched with other RDF triples to offer a richer cultural heritage experience. Linked open data releases each museum, archive, or library's collection from its silo, allowing it to be placed in the context of similar cultural heritage collections and knowledge sets from around the world. We reviewed projects that exemplified different levels of openness and linking. Many projects are still in conceptual phases, and others stop at certain stages of development after running out of energy or funding. We aimed to illustrate the best of linked data projects while acknowledging that no one project is outstanding in every aspect. In some cases a linked open data project embodied the spirit of openness, but may not have met every technical specification. This paper attempts to survey the landscape of linked open data projects by placing fifteen projects within the context of a linked data lifecycle and highlighting areas in which each project excels.



Figure 2: Six Stage LOD project Life Cycle

Methodology

Each project was initially evaluated on the following criteria:

- Affiliation/Mission/Intended Audience
- Knowledge Org/Data Models & Vocab and/or Technology Platform
- Usability/Interface Design
- Discovery (search & nav)
- Data Shareability (ie. availability of an API)
- Sustainability (i.e., digital preservation, documentation, or available code)
- Project Leaders
- Funding Sources
- Level of Collaboration
- Analysis
- Star-Rating (i.e., Tim Berners-Lee's coffee cup)

To allow for easy comparison, our initial survey and evaluation results were placed in a spreadsheet using these criteria as categories. After getting a sense of the linked open data landscape, we selected fifteen projects to include in our review. These projects demonstrated innovative thinking or exemplary product in at least one stage of the linked open data cycle.

Tim Berners-Lee's Five Star Model for linked open data is a standard for the movement, but we found it limited when discussing projects that were innovative in some aspect but ran out of money before finishing, or had developed a prototype as a proof of concept but never saw a need to develop a working model. The Five Star Model is valuable when describing the openness and linking potential of the data, but not when describing what cultural heritage developers are doing with the data. Many projects had achieved their goal of creating an open dataset or linking it to the greater cultural heritage field, but had never developed a point of access for users. Yet all these projects seemed to contribute something valuable to the linked open data field.

We adopted the six-stage model to reflect the realities of grant-based project development, and recognize the accomplishments of each cultural heritage team. Each of these projects shows initiative in at least one aspect of linked open data as originally presented by Berners-Lee, and we chose to recognize those accomplishments without placing a star-based value judgment on their work.

Projects

1 Stage 1: Developing Datasets

The first stage to a successful linked open data project is releasing a dataset that is usable for other projects. Ideally datasets are converted into RDF triples and shared via an open API or SPARQL query endpoint.

Here are a few examples of cultural heritage institutes that developed data well. First we look at an organization that saw releasing their catalog in linked open data as a way to maintain and grow its influence in the world of cataloging.

1.1 The Library of Congress

1.1.1 Introduction

As part of an initiative to drive American libraries and scholarly institutions toward sharing data more openly, the Library of Congress has embarked on a Bibliographic Framework Plan. This plan will encourage libraries as well as the Library of Congress itself to slowly transition from MARC records toward RDA and linked open data. Since 2009, the Library of Congress has been exposing its famous vocabulary and its Authority Names into linked open data through the Library of Congress Linked Open Data Service portal. The goals of the Library of Congress Linked Open Data Service are twofold, benefiting both the Library of Congress and human and machine users.

Benefits for the Library of Congress include the ability to establish provenance for the loc.gov domain on the Semantic Web, fewer taxing downloads from its website related to users retrieving single records at a time, and finally the ability to share best practices and encourage other institutions to format their data as linked open data. The benefits for human and computer users are also significant and include bulk download of entire controlled vocabulary datasets, the ability to link to Library of Congress data values within their own metadata via linked data, access to powerful datasets at no cost, concept and value relationship mapping, and a host of other developer-specific services.

1.1.2 Authorities and Vocabularies

The Library of Congress Authorities and Subject Headings are used throughout the world to resolve name variation and to aid in cataloging and research. The Library of Congress has been regarded as a trusted authority since the 1800s; the Library is now continuing this

legacy by coining authoritative and reliable URIs based on its existing vocabularies and authorities. Due to the large scale of this project and the incredible amount of data to be formatted, the Library of Congress will be releasing linked data versions of different Library of Congress Classification classes as they are finished. The Linked Data program launched with the release of the B (Philosophy, Psychology, and Religion), M (Music), N (Fine Arts), and Z (Bibliography, Library Science, and Information Resources) classes.

By visiting the Library of Congress Linked Data Service website (<http://id.loc.gov>), users can query any of the available datasets as well as any combination of the datasets simultaneously. The Linked Data Service search tool works similarly to the traditional Authorities portal but includes more information as well as an updated look and simplified results pages. Searching under a related name (e.g., Lady Day for Billie Holiday), users are taken directly to the authority file where popular Library of Congress information can be found; the files URI, links to alternative formats, and exact matching concepts from other schemes are also provided.

The power of the Linked Data Services search portal and the Authority and Vocabulary file pages is the elegant connections each file forms from well-known Library of Congress services such as LC Classification and source links to linked open data and through LOD to other schemes and thereby other reputable institutions.

1.1.3 Datasets and Documentation

The Library of Congress currently provides seventeen downloadable datasets to the public. The datasets are formatted in RDF/XML, Turtle, or N-Triples; each download also contains MADS/RDF and SKOS/RDF representations of the data. APIs are not available through the Library of Congress.

The Library provides instruction, explanation and definitions through its Technical Center (<http://id.loc.gov/techcenter/>), which walks users through downloading and utilizing the bulk dataset downloads and includes a brief guide through command-line downloads. The Technical Center also helps to promote the Library's focus on education and the dissemination of LOD through explanatory pages devoted to metadata, searching/querying, and serializations. The teaching and promotion efforts are extended to a wider audience through blogposts from the Library of Congress website that point to the Linked Data Services webpage.

1.2 Out of the Trenches: Linked Open Data of the First World War

1.2.1 Introduction

This project demonstrates the collaborative nature of many LOD projects. Here, multiple cultural heritage groups joined to release collections as linked open data.

The Pan-Canadian Documentary Heritage Network (PCDHN) is a collaboration of libraries, archives, and museums (LAMs) formed at an initial stakeholders forum held in October 2010. The network was envisioned as a network of networks that would address the increasing volume of information in the face of dwindling resources and growing demand for access to content. Participants at its second forum (May 2011) proposed several goals, including the collaborative acquisition, management, preservation and access to relevant, coherent and sustainable documentary heritage resources, and decided on five initiatives, including a pilot project on metadata and resource discovery (LAC, 2012).

The pilot project developed into a proof-of-concept to showcase a sample of the networks resources using linked open data. The partners in the project—Library and Archives Canada (LAC), Canadiana, Bibliothèque et Archives nationales de Québec (BAnQ), University of Calgary Libraries and Cultural Resources, University of Saskatchewan Library, University of Alberta Libraries, McGill University Library, Universit de Montral, Canadian Museum for Human Rights, Canadian Historical Association, and Saskatchewan Archives Board—selected World War I as the topic, naming the project Out of the Trenches. The results include a project final report, a demonstration video, and the RDF metadata, all of which are hosted by Canadiana (<http://www.canadiana.ca>), a membership alliance dedicated to building Canadas digital preservation infrastructure and providing the broadest possible access to Canadian documentary heritage. Canadiana works to preserve Canadas print history and make it accessible online; its Early Canadiana Online service includes 4 million pages of published heritage materials.

The PCDHN documentation was published on the Canadiana site in July 2012; no additional news postings, follow up, or response have yet to appear, although the reports recommendations include further development of visualizations, refinement of query results, and the addition of resources and metadata (PCDHN, 2012, p. 17).

1.2.2 The Project

According to the project report, the major reasons for the adoption of LOD are providing access, enabling the reuse of data and metadata, connecting resources from a broad range of institutions to create a richer experience for the public, and allowing LAMs to showcase their resources across organizational and subject boundaries. The goals are to enhance

discovery and make collections more accessible and meaningful.

Resources, with the data submitted in various formats, were contributed by five of the partners, as follows:

- BAnQ: war songs (MARC records)
- McGill: war posters (spreadsheets)
- U Alberta archives of newspaper articles, postcards, and wartime records (MODS XML)
- U Calgary: portrait archives of Canadian Expeditionary Force soldiers and World War I documents (spreadsheets)
- U Saskatchewan: archival materials from the Saskatchewan War Experience Project (DC RDF)

The material was then expressed as RDF triples. These along with existing resources, such as Geonames, VIAF (the Virtual International Authority File, one of the largest LOD datasets; see <http://viaf.org>), and Canadian Name Authorities, were used to create PCDHN authorities. The project also used URIs from published authorities, including LCSH, TGM, Rameau (Bibliothèque nationale de France), and LAC Canadian Subject Headings. The project report includes an appendix showing the rules used to map the metadata from the format in which it was submitted to the selected ontologies.

1.2.3 Metadata Model

The project created a metadata model for the contributed resources, informed by existing models such as FRBR, CICOC, and ISAD/ISAAR).

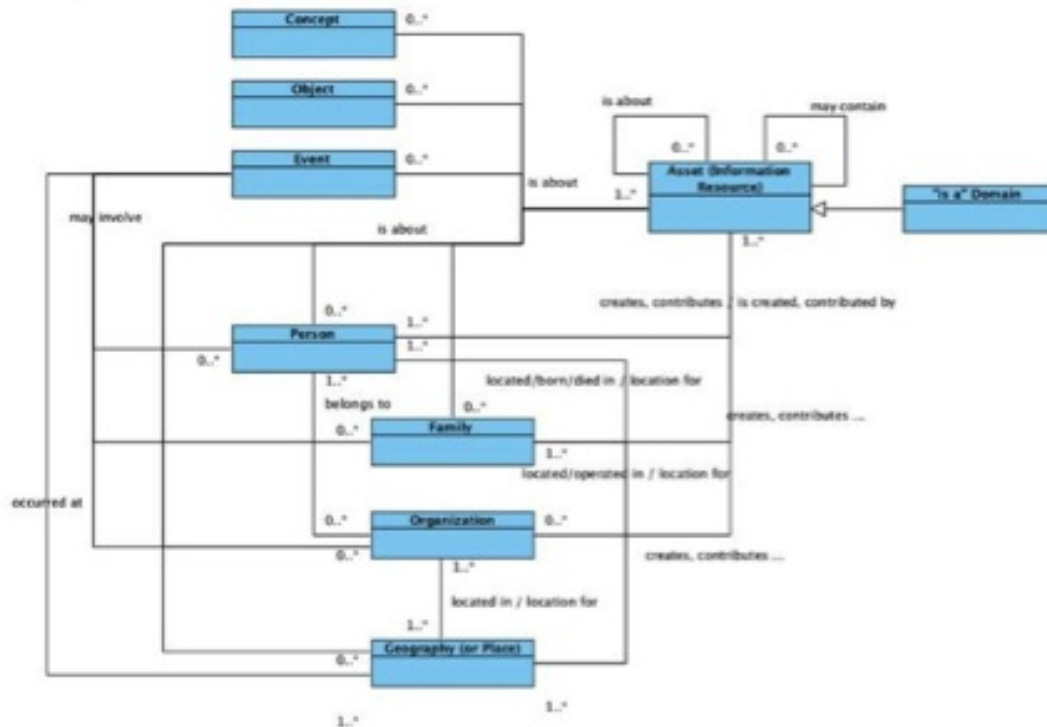


Figure 3: Out of the Trenches Metadata Model (PCDHN, 2012, p. 6)

The metadata model was used to guide the mapping of the contributed metadata to existing/published RDF ontologies and element sets. The following ontologies / element sets were used in the metadata mapping:

- Bibliographic Ontology
- Dublin Core Ontology
- Biographical Ontology
- Event Ontology
- ChangeSet Ontology

- Friend-of-a-Friend Ontology
- Open Archives Initiative Object Reuse and Exchange
- IFLA International Standard Bibliographic Description (ISBD) Ontology
- Resource Discovery and Access (RDA) Element Sets
- Entities (Classes)
- Work, Expression, Manifestation, and Item (WEMI) properties
- Person, Family, and Corporate Body properties
- Concept, Event, Object, and Place properties
- Role properties
- Relationship properties for WEMI
- Relationship properties for Person, Family, Corporate Body
- Relationship properties for Concept, Event, Object, Place
- W3C Ontologies for RDF/RDF Schema, OWL (base ontologies), time, and Simple Knowledge Organization System (SKOS)
- e-mail Message Ontology

(PCDHN, 2012, pp. 6-7)

1.2.4 Technologies and Metadata

The proof-of-concept uses the following technologies:

- Mulgara triple-store: for storing the data statements for the resources and providing a SPARQL endpoint for querying the data store (hosted by the University of Alberta and locally);
- Javascript in a browser (Firefox only for the purpose of the proof-of-concept): for querying the triple-store and the actual visualization of the results of the query;
- Web server: for storing cached resources (such as, cached pages of the university website pages for individual resources, vignettes of soldiers for the exclusive use by the application, LAC resources cached for performance reasons).

(PCDHN, 2012, p. 13)

Note that though mentioned, the SPARQL endpoint does not appear to be available to the public.

The dataset for each ontology was then released as raw RDF/XML files. As an example, a sample of the Events Ontology metadata is shown below.

```
-<rdf:RDF>
-  <rdf:Description rdf:about="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONLPCDHN0001-EventResourceMap">
    <rdf:type rdf:resource="http://www.openarchives.org/ore/terms/ResourceMap"/>
    <ore:describes rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONLPCDHN0001-EventResource"/>
  -  <dc:provenance>
    -  <cs:ChangeSet>
      <cs:subjectOfChange rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONLPCDHN0001-EventResourceMap"/>
      <cs:creatorName rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/organization/PCDHN0058"/>
      <cs:createdDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2012-04-15T00:00:00</cs:createdDate>
      <cs:changeReason xml:lang="en" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">created</cs:changeReason>
      <cs:changeReason xml:lang="fr" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">créé</cs:changeReason>
    -  </cs:ChangeSet>
    </dc:provenance>
  -  <rdf:Description>
-  <rdf:Description rdf:about="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONLPCDHN0001-EventResource">
    <rdf:type rdf:resource="http://www.openarchives.org/ore/terms/Aggregation"/>
    <ore:isDescribedBy rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONLPCDHN0001-EventResourceMap"/>
  -  <dc:title xml:lang="en">
    Library and Archives Canada - Resources on the Battle of Passchendaele
    </dc:title>
  -  <dc:title xml:lang="fr">
    Bibliothèque et Archives Canada - Ressources sur la bataille de Passchendaele
    </dc:title>
    <!-- World War, 1914-1918 CHIs -->
    <dc:subject rdf:resource="http://experimental.worldcat.org/Est/1180746"/>
    <dc:subject rdf:resource="http://id.loc.gov/authorities/subjects/sh85148236"/>
    <dc:subject rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/RVMConcept/RVM001"/>
    <dc:subject rdf:resource="http://www.collectionscanada.gc.ca/900F11/040004/CaOONL0200E0623"/>
    <dc:subject rdf:resource="http://www.collectionscanada.gc.ca/900F11/040004/CaOONL1002E1618"/>
    <!-- Canada -->
    <dc:subject rdf:resource="http://www.geonames.org/6251999"/>
    <!-- Subjects common to all aggregated resources -->
    <dc:subject rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/PCDHN0001"/>
    <dc:subject rdf:resource="http://www.geonames.org/2789270"/>
    <!-- Aggregated resources -->
    <ore:aggregates rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONL-a040139"/>
    <ore:aggregates rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONL-a040138-v6"/>
    <ore:aggregates rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONL-a008293261-v8"/>
    <ore:aggregates rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONL-a008293265-v8"/>
    <ore:aggregates rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONLPCDHN0001-EventArt"/>
    <ore:aggregates rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONLPCDHN0001-EventMonument"/>
  -  <dc:provenance>
    -  <cs:ChangeSet>
      <cs:subjectOfChange rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/event/resource/CaOONLPCDHN0001-EventResource"/>
      <cs:creatorName rdf:resource="http://rdf.canadiana.ca/PCDHN-LOD/organization/PCDHN0058"/>
      <cs:createdDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2012-04-15T00:00:00</cs:createdDate>
      <cs:changeReason xml:lang="en" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">created</cs:changeReason>
      <cs:changeReason xml:lang="fr" rdf:datatype="http://www.w3.org/2001/XMLSchema#string">créé</cs:changeReason>
    -  </cs:ChangeSet>
    </dc:provenance>
  -  <rdf:Description>
```

Figure 4: Sample of Out of the Trenches Events Ontology RDF/XML (pcdhn-lod.zip/Complete Files/EventResources.rdf)

1.2.5 Analysis

In publishing its datasets under an open license, using URIs to identify items, and linking to other datasets such as Library of Congress Subject Headings and Geonames, the PCDHN has fulfilled the goals of the five-star open data model. The lack of public access to the SPARQL query endpoint, however, makes the datasets less likely to be used. Given the current high level of interest among LAMs to create projects focused on World War I due to the upcoming centennial of the wars outbreak in 2014, an effort to coordinate and combine these projects would be a valuable opportunity to test LODs potential.

As noted above, the PCDHN also created a proof-of-concept to demonstrate how the data could be used, presenting its visualization in a brief video. This phase of the project is discussed in Section 4, Developing User Interfaces, below.

2 Stage 2: Linking Data

Once linked open datasets are available, it is up to cultural heritage institutions to determine how to use them to enrich their own collections. Libraries, museums, and archives can offer their patrons additional context for understanding their collections by integrating other linked open datasets into their websites and apps and encouraging patrons to make new connections.

We begin this section with an institute that was an early forerunner in LOD, which released its dataset early on. Rather than augment its own collections with other linked open datasets, it chose to integrate its catalog with the rapidly growing European Open Access Library (<http://www.oapen.org>).

2.1 The Hungarian National Széchényi Library

2.1.1 Introduction

Officially announced in April 2010, the Hungarian National Library’s shared catalog was one of the first successful LOD projects. Using practices previously set forth by the Swedish LIBRIS and the tutorial “How to Publish Linked Open Data on the Web”, National Széchényi Library released their OPAC, Digital Library and authority data as LOD. Led by Adám Horváth NSZL worked to configure their catalogue in a web context with tenets set down by Anders Söderbäck of the National Library of Sweden. The goals were:

- Need to be open (Get your data out!)
- Need to be linkable
- Needs to provide links
- Must be part of the network
- Can not be an end in itself
- Allow for hackability

NSZL was involved in grant funded work collaborating with the European Open Library. This work has been completed.

2.1.2 Authorities and Vocabularies

The NSZL links incorporate RDFDC for bibliographic data, FOAF for name authorities, SKOS for subject and geographical terms, DBpedia name files, CoolURIs, owl:sameAS

statements and is available via a SPARQL endpoint. The Hungarian National Szchnyi Library uses a Virtuoso SPARQL endpoint and Jena RDF database to make the contents of their OPAC, Digital Library and authority data available as Linked Open Data to initiatives like the European Open Library where Hungarian Library holdings can be seen in context with similar items from items in national libraries across Europe. The library links incorporate RDFDC for bibliographic data, FOAF for name authorities, SKOS for subject and geographical terms, DBpedia name files, CoolURIs and owl:sameAS statements in order to allow for discoverability and broad connections over the array of data.

2.1.3 Analysis

The Hungarian National Szèchènyi Library completed a project showing that Linked Open Data is indeed an attainable goal and has now moved on to other things, such as a project on e-books. This was the first and perhaps the most important step for LOD. Another important step for LOD is documentation. In order to maintain sharability potential collaborators will need to know how LOD works and one good way to learn this is by studying successful projects. Since the completion of the LOD project the information available about it on the NSZL website has dwindled. Some links are broken and the bulk of the information is available on a section called “semantic web”. Using current LOD terminology would help with the discoverability of the project. This lends itself to the point that although LOD projects throughout Europe are making strides the nuances of language and communicating in multiple languages is at times still an issue. NSZL may look simplistic in the current life cycle but it was equally as ambitious as some of the later projects when it was a first. Later projects could not have branched out in the way they have with out the basic groundwork provided by NSZL.

2.2 Civil War Data 150: Unfinished Dreams

2.2.1 Introduction

Headed by Jon Voss, Civil War Data 150 initially began as a partnership between the Archives of Michigan, the Internet Archive, and Freebase. The project later added two additional partners: the University of Richmond and HistoryPin—Voss's newest project. Civil War Data 150 (CWD150) was created to celebrate the sesquicentennial of the American Civil War, beginning in April 2011. The project's initial plans included continued growth of the project over the four years of the sesquicentennial, from 2011 to 2015. CWD150 was a multifaceted project that encompassed—and planned to encompass—a number of different data sources, tools, and applications as well as a social media component. Along with promoting the digitization of archival documents from the Civil War, CWD150 championed the adoption of LOD. Libraries, archives, museums, and even individual researchers were encouraged to contribute Civil War data to the project.

CWD 150's stated goals in a 2010 presentation were:

1. Identify sources and map metadata into Freebase.
2. Create Web apps to enable users to add to or modify shared metadata with strong identifiers.
3. Engage the public in the process of interacting with and adding value to the data.

(civilwardata150.net)

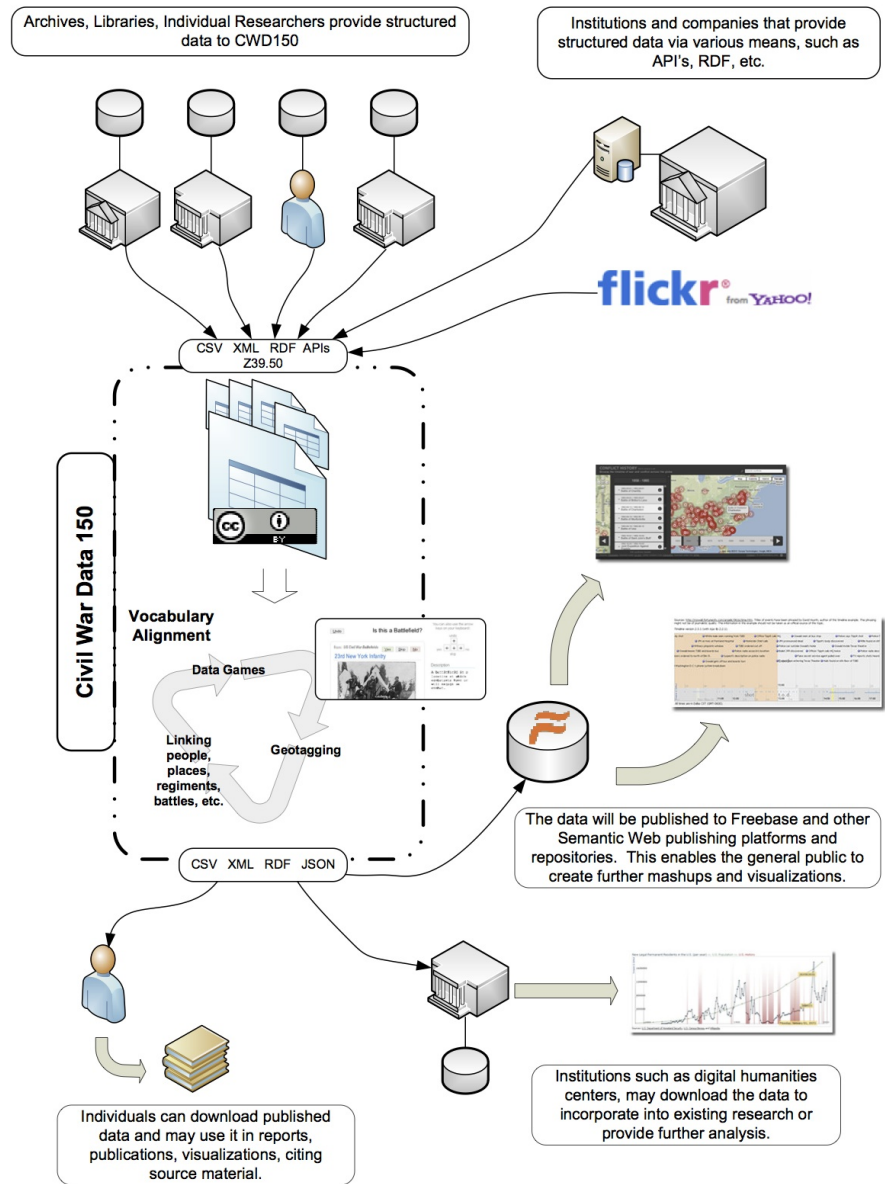


Figure 5: Civil War Data 150 Workflow

(Jon Voss, 2010)

2.2.2 Data and Technology

Although, as noted above, CWD 150 had hoped to gather data from archives, libraries, museums, and individuals throughout the world, the majority of the data used in the finished applications came from Wikipedia (via Freebase) and Tuft Universitys Perseus Digital Library (www.perseus.tufts.edu/). CWD 150 planned to gather data not only through submissions and partnerships, but also to identify primary datasets through Web crawling, screen scraping, XML sums, and CSV files. The next step in the project was to map the metadata in Freebase. Once the data was structured, CWD 150 planned to create its own Web apps and make its data stronger by encouraging Civil War historians and others to add specific schema through a crowdsourcing tool.

Web Applications and Tools

2.2.3 Conflict History

ConflictHistory.com was a Web application that used the Freebase and Google Maps APIs to create a dynamic, geographic timeline of the history of war and conflict throughout the world. The majority of the text utilized by the Freebase API came from Wikipedia; users of Conflict History were encouraged to edit the information on either Freebase or Wikipedia. Conflict History was developed by TecToys, a Web and mobile studio founded by Thimon Sistermanns. In October 2012, Sistermanns announced on the Conflict History website that the application would be taken down to be updated as the data sources and Web technologies tool employed were quickly becoming outdated. Although the announcement stated that the site would be relaunched in January 2013, as of March 2013 this notice is still posted on the website and the tool does not appear to be updated.

As they seem to have been two distinct projects, CWD150's involvement with Conflict History may have been more as a supportive player than a direct partner. Although the two projects do not appear to have collaborated, the benefits of supporting and promoting one-another definitely could have led to improvements in both. CWD 150s call to action toward the releasing of Civil War data as linked data would have immediately enhanced Conflict History, as the Freebase API would have benefited from these updates, therefore making Conflict History a more powerful tool. In turn, Conflict History would have benefited CWD 150 through its dynamic timeline, since CWD 150 could have encouraged the release of data for it without having to devote time and resources to creating its own mashup.

2.2.4 Live Tweeting the Civil War + 150

Live Tweeting the Civil War + 150 is a now-defunct project of CWD 150. The project used Frederick Dyers 1908 Compendium of the War of the Rebellion, which was turned into structured data by the Perseus Project. Unfortunately, the downloadable XML from Perseus was not structured as LOD, which CWD 150 planned to remedy by converting the XML to CSV so that it could be placed in a relational database and broken down by event date. Next CWD 150 planned to use a Python script—inspired by Ed Summers PaperBot 7—to automatically create tweets appropriate to each day of the sesquicentennial. Finally CWD 150 hoped to add a link to each tweet directing users to different resources.

2.2.5 Analysis, or What Can Be Learned from a Promising, Unfinished Project

Although CWD 150 is an unfinished project, there is still much to be learned from the documentation surrounding the project through blogs, presentations, podcasts, and articles. Primarily, CWD 150 is a significant project in the short history of LOD, as it created a call to arms for the creation of LOD and the sharing of archival documents and materials throughout the country. The issues that led to the termination of the project are not stated anywhere on the projects website, but it can be assumed that this ambitious project did not have the staffing necessary to fulfill all of its goals.

2.3 Linking Lives

This project demonstrates some of the pitfalls of linking data. In this project, not all institutions had the same quality of data, and the developers had to compensate once the linked open datasets were brought together.

2.3.1 Background

Linking Lives is an LOD test project focusing on biographical information using linked data from the Archives Hub, based at Mimas at the University of Manchester in the UK, and external datasets (DBpedia, VIAF, etc.). Linking Lives builds on, and is an extension of, the LOCAH (Linked Open Copac & Archives Hub) project of 2010-2011, and was also a one-year project that ran from 2011 to 2012. Funding was provided by the JISC (which originally stood for Joint Information Systems Committee), a charitable organization working on behalf of UK education to champion the use of digital technologies. By linking to larger datasets, Linking Lives creates the potential for a richer set of results for those searching for biographical information, especially in bringing to light relevant materials deep within archival collections that researchers might otherwise not discover.

2.3.2 Archives Hub

The Archives Hub (<http://archiveshub.ac.uk>) represents more than 220 institutions across the UK. In effect, it is a collection of collections, bringing together diverse resources to create opportunities for scholarly research. Its objectives include facilitating and expanding access to digital content, developing sustainable methods for data ingest, and developing functionality and features to facilitate the use and maximize the benefits of the data. LOCAH and Linking Lives are efforts to explore the possibilities of this last objective.

The Archives Hub is intended to be interoperable with other services and systems. Descriptions are stored as EAD and transformed into HTML in order to be viewed in a browser. The Hub supports the SRU and Z39.50 search protocols and OAI-PMH (used to harvest descriptions) as machine-to-machine interfaces. Descriptions are created according to ISAD(G), the international standard for finding aids. Index terms also must follow recognized rules or be taken from recognized sources (e.g., NCA Rules, AACR2, UKAT), but the Hub does not prescribe which rules/sources should be used, as long as the provider specifies which set of rules or sources are being employed.

2.3.3 LOCAH

The LOCAH project (<http://archiveshub.ac.uk/locah>) was an effort to publish data from a sample of Archives Hub finding aids and from Copac (another Mimas service funding by JISC, which brings together the catalogs of more than 70 major UK and Irish libraries, including the British Library, many university libraries, and specialist research libraries) as linked open data. The following resources are available:

- SPARQL endpoint: <http://data.archiveshub.ac.uk/sparql>
- Query box for trying out SPARQL queries: <http://data.archiveshub.ac.uk:8080/test/>
- RDF dump of the dataset: <http://data.archiveshub.ac.uk/dump/linkedarchiveshub.zip>
- Archives Hub EAD to RDF XSLT stylesheet: <http://data.archiveshub.ac.uk/ead2rdf/>

An example of the front-end interface for the linked data is provided for the Beverly Skinner Collection (see screenshot); note especially the `associatedWith` property, which includes values based on NCA rules, AACR2, and Unesco datasets.

Property	Value
archiveshub:accessProvideBy	• <http://data.archiveshub.ac.uk/id/repository/gb1086>
archiveshub:accessRestrictions	• Access for visitors is by appointment only. (EN)
archiveshub:associatedURL	• <http://data.archiveshub.ac.uk/id/concept/organisation/charities/hammersmitharchive> • <http://data.archiveshub.ac.uk/id/concept/person/actors/skinnerbeverley1938-1999artist> • <http://data.archiveshub.ac.uk/id/concept/place/london/england> • <http://data.archiveshub.ac.uk/id/concept/place/london/england> • <http://data.archiveshub.ac.uk/id/concept/thesauri/occupation> • <http://data.archiveshub.ac.uk/id/concept/thesauri/womenartists>
dc:time:date	• 1966-1990
archiveshub:dateCreatedAccumulatedEnd	• 1990 (iso8601Year)
archiveshub:dateCreatedAccumulatedStart	• 1966 (iso8601Year)
archiveshub:dateCreatedAccumulatedString	• 1966-1990
dc:time:extent	• <http://data.archiveshub.ac.uk/id/extent/gb-1086-gb1086skinner>
archiveshub:extent	• 1 box and 15 slides
archiveshub:hasBiographicalHistory	• <http://data.archiveshub.ac.uk/id/biography/gb-1086-gb1086skinner>
dc:time:identifier	• GB 1086 Skinner
ids:label	• Beverly Skinner Collection (EN)
dc:time:language	• <http://lexo.org/id/iso639-3/eng> • <http://lexo.org/id/iso639-3/eng>
archiveshub:level	• <http://data.archiveshub.ac.uk/id/level/fonds>
archiveshub:origination	• <http://data.archiveshub.ac.uk/id/origin/gb1086-beverleyskinner1938-1999>
archiveshub:otherFindingAids	• List available at the Library (database) (EN)
foaf:page	• <http://archiveshub.ac.uk/data/gb1086skinner>
skos:prefLabel	• Beverly Skinner Collection (EN)
archiveshub:processing	• Description by Althea Greenan, MAE 2002. Submitted to the Archives Hub as part of Genesis 2009 Project. (EN)
event:produced_in	• <http://data.archiveshub.ac.uk/id/production/gb-1086-gb1086skinner>
archiveshub:scopecontent	• The collection consists of newsclippings of exhibitions including 'Woman Stage', Bristol, 1972 which highlighted the sp...
ids:seeAlso	• <http://archiveshub.ac.uk/data/gb1086skinner> • <http://data.archiveshub.ac.uk/id/archivalresource/gb1086skinner>
dc:time:title	• Beverly Skinner Collection (EN)
ids:type	• archiveshub:ArchivalResource

This page shows information obtained from the SPARQL endpoint at <http://data.archiveshub.ac.uk/sparql>
[As Text](#) | [As RDF/XML](#) | [Browse in Disco](#) | [Browse in Tabulator](#) | [Browse in Conet for Browser](#)

Figure 6: Screenshot of Archives Hub/LOCAH front-end interface for Beverly Skinner collection

(<http://data.archiveshub.ac.uk/page/archivalresource/gb-1086-gb1086skinner>)

The LOCAH Linked Archives Hub data and content are licensed under a Creative Commons CCO 1.0 license. Participants in the LOCAH project were Mimas, UKOLN, Eduserv, Talis, and OCLC Research. The project team included Adrian Stevenson of UKOLN; Jane Stevenson, Archives Hub Coordinator at Mimas; Pete Johnson of Eduserv, and others. OCLC's involvement was mainly to help out with VIAF.

In a brief feature article on the project, Adrian Stevenson describes its value as allowing the development of new channels into the data. Researchers are more likely to discover sources that may materially affect their research outcomes, and the hidden collections of archives and special collections are more likely to be exposed and used (Ruddock & Stevenson, 2011, p. 19). This article notes that the major challenge is variance in the data; although the libraries and archives providing the data adhered to standards, these standards are in practice only guidelines that get interpreted and implemented in different ways at the local

level, which affects the ability to machine-process information.

With the datasets available as XML, the transformation was not difficult to carry out. The archival data were transformed using an XSLT; the library data with Java. In addition to the SPARQL endpoint, the data may be downloaded in RDF/XML, JSON, and Turtle. The project also created external links to other datasets, such as VIAF and LCSH, using same as relationships.

2.3.4 Linking Lives

Linking Lives (<http://archiveshub.ac.uk/linkinglives/>) expanded on LOCAH by bringing in more external datasets and creating a model of a Web interface for researchers that presents name-based biographical pages that include information about archives relating to individuals alongside other resources, including Wikipedia (via DBpedia). This interface does not exist, but is merely mocked-up and described in an article by a team member (Stevenson, 2012). Linking Lives is a proof of concept to show how the LOD of the LOCAH project could be used and how it could be beneficial to researchers.

Linking Lives focuses on individuals as a way into archival collections as well as other relevant data sources (Stevenson, 2012, p. 15). One goal was to expose archival collections to researchers, who might not be familiar with primary sources or who might not think of searching archival collections when starting biographical research.

A mock-up of an interface was developed (see screenshot). One noteworthy feature is that the content boxes allow the user to clearly distinguish among the data sources, allowing decisions to be made based on provenance and trust of the data source. Another valuable feature is a tab to list alternative names drawn from various sources, including VIAF.

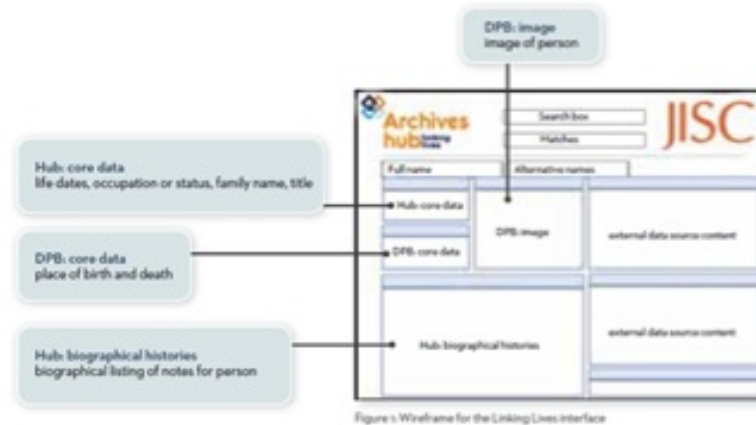


Figure 7: Screenshot of Linking Lives interface wireframe showing content boxes that identify sources of data

(Stevenson, 2012, p. 15)

As with LOCAH, data inconsistency is one of the largest obstacles. Stevenson (2012, p. 16) gives as an example even the very basic issue of the identifiers for the source archives themselves: since the primary role of this information in the Archives Hub was to be a displayed visual indicator to end-users, some inconsistency in the structure of these descriptive codes was not a problem. In this situation, rather than attempting to work around the problem, the best solution was to go back in and rectify the information. Another example of inconsistency that would commonly occur with archive-based data projects is in the description of extent; since this may be expressed as a number of boxes, a measure of linear space (meters, feet), or a text-based description (e.g., photographs and drawings in 3 outsized boxes [Stevenson, 2012, p. 16]).

External datasets the Linking Lives team looked at include DBpedia, OpenLibrary, VIAF, Freebase, BBC Programmes, and Linked Open British National Biography. The effort to add more links to external data exposed several issues, including lack of control over the data, sources that do not provide all of their data as linked data (Stevenson [2012, p. 17] notes that even VIAF does this in not including its lists of selected titles for authors as linked data), and sources that do not provided a SPARQL endpoint. Most interesting, the Linking Lives project team struggled to find linked data that includes connections between people; for example, a simple statement that x person knows y person (Stevenson, 2012, p. 17; cf. the centrality of this relationship to the Linked Jazz project). Another problem is where desired data are not always available, for example an image of an individual from DBpedia.

Another problem the project encountered is in resolving ambiguities in the identification of individuals. This occurs both in instances where the name format and/or description of the same individual differ across datasets and where the data matches exactly, but describes different individuals with the same name. The Linking Lives team expended considerable time and energy in manual rectification of name information, which presents an obstacle that other projects would have to take into consideration when planning and allocating resources. The necessity of performing this sort of clean-up operation, suggests that linking to larger datasets is only a first step in what may be an arduous, labor-intensive process.

Jane Stevenson states in conclusion, the biggest single factor in terms of additional work has been clearing up our own data and that it needs to be easier to locate and probe data sources . . . [and] to link to these external sources. . . . A degree of investigation may be necessary to establish exactly what is being provided and how uniform it is (Stevenson, 2012, p. 22). She then notes a different sort of problem. Despite continued expansion of the LODLAM community and interest in LOD, it may be evolving too slowly to attract the business investment necessary to make it viable. Does the altruistic goal of opening up data to advance knowledge and benefit research provide a strong enough impetus to drive the linked data ideal? (Stevenson, 2012, p. 23).

2.3.5 Analysis

As with other LOD projects described in this paper, sustainability is a significant issue. Having been produced based on two separate one-year grants, both now completed, there is no indication that any further work will be done on the Linking Lives project. Will the model interface be completed and be made accessible to users? If it were available, are general users, as opposed to specialist researchers, likely to find and use the Linking Lives site when searching for biographical information? Will the linked open data be used in other ways, either by the project team or others? Or does the project simply serve as a test case of what is possible? Follow-up reporting from the project team would provide a helpful gauge of the ongoing value of creating LOD-based projects.

2.4 Bibliothèque nationale de France

2.4.1 Introduction

Established in the sixteenth century, the Bibliothèque nationale de France (BNF) is comprised of more than fourteen divisions and remains the largest cultural heritage institution in France today. BNF has strong financial support from the French government, and there are even cultural public policies in place that advocate making culture and art accessible to the public. This movement to release public data has been adopted by the French administration since 2010. In 2011 the BNF launched its rich linked data resource, becoming an LOD service. Its mission is to collect, conserve, and share the documentary heritage of France.

2.4.2 Project Data

With the launch of its linked open data project, called data.bnf.fr, BNF used both conceptual models for the organization of bibliographic information, such as FRBR, and linked data technologies to provide integrated access to both BNF's collection and related resources available on the Web. BNF released two sets of data for this project; digitized documents and metadata. Data.BNF.FR allows data to be extracted, transformed, and consolidated into a common database, and allows interoperability by providing URIs for resources and ARK identifiers of the record or scanned document, as shown below.

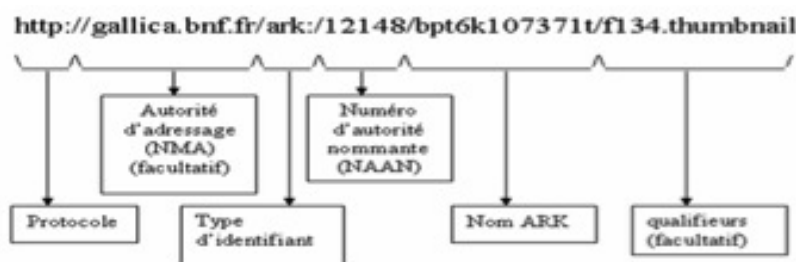


Figure 8: ARK Identifier Structure
(Bibliothèque nationale de France, 2013)

2.4.3 Main Aspects and Tools

BNF uses technical standards in compliance with the Semantic Web and uses linked data in line with W3 open licensing, which allows free reuse, including commercial use, with attribution to the owner. BNF uses URIs for all its resources; each resource has a permanent

identifier, known as an ARK identifier. The BNF identifies bibliographic and authority records with descriptions and digital ARK identifiers. These ARK identifiers help create the linkages of general items within their respective catalogs, including BNF's digital catalog Gallica. Each page within the entire database Data.bnf.fr contains RDFs, which in turn provide linked open data. BNF follows the standards for bibliographic description and identifies authority records to describe people, organizations, works, and themes. Finally, BNF uses alignment techniques and data federation.

2.4.4 Data Extraction and Presentation

The data.bnf.fr project extracts data generated in various formats, including InterMarc for book catalogs, XML-EAD for archival inventories, and DC Digital Library. The data are modeled and grouped by automatic processing and published according to the standards of the Semantic Web. The data are then presented via <http://data.bnf.fr/semanticweb> according to several syntaxes such as RDF or RDF/XML, N3, RDF-NT, and JSON. Data.bnf.fr aligns its datasets with institutions and schemas such as the Library of Congress, Dewey Decimal System, DBpedia, Virtual International Authority File, Dublin Core Metadata Initiative, and CKAN.

2.4.5 Ontologies and Vocabularies

To promote interoperability, BNF reuses existing vocabularies such as RDF, DC, FOAF, RDAgroup2elements, RDA relationships, RD Vocab, OWL, and OAI-ORE (Open Archives Initiative Object Reuse and Exchange). BNF uses some controlled vocabularies provided by the LOC, but some of BNF's vocabularies cannot be expressed as ontologies, as they are specific to meet the needs of the institution. These vocabularies produce bibliographic records in MARC format, InterMarc, and are available in HTML and RDF. The three main vocabularies at BNF are related to country codes, relater codes, and RAMEAU subject headings.

2.4.6 FRBR Concepts and RDF Data Model

Data.bnf.fr. has adapted and also experimented with the FRBR model, by linking the entity groups together through relationships: information about documents, persons and organizations, and subjects. The first group (work) of the FRBR model describes the different aspects of an intellectual or art creation, and discerns four levels: work, expression, manifestation, and item. The second group ("author") describes a person or organization that can be identified as an author of a work or a contributor to a work. All events by a single author are attributed to "works" inside the original bibliographic record. Third,

there are “subject” records from BNF, which RAMEAU is indexing. These have been converted into the RDF language SKOS (Simple Knowledge Organization). Some subjects are not linked to the title authority record, and remain orphaned.

BNF notes that its model is not completely FRBRized; therefore it aims to improve the way its data is translated into FRBR by aligning or clustering these orphaned subjects. Through the alignment proposal, subjects would be aligned to a work, both of which would have their own title and authority record. These events would not have any link to the title authority record, and would come from the bibliographic records of the main catalog or from descriptions of BNF archives and manuscripts. BNF proposes using a simple yet advanced alignment algorithm to determine whether the two character strings correspond to the same work. The link to the author record remains essential to align the works. Clustering would work if there is no title authority record, and where subjects are attributed to a new unit.

The two diagrams below shows the general organization of data in data.bnf.fr, based on the main entities of the FRBR model:

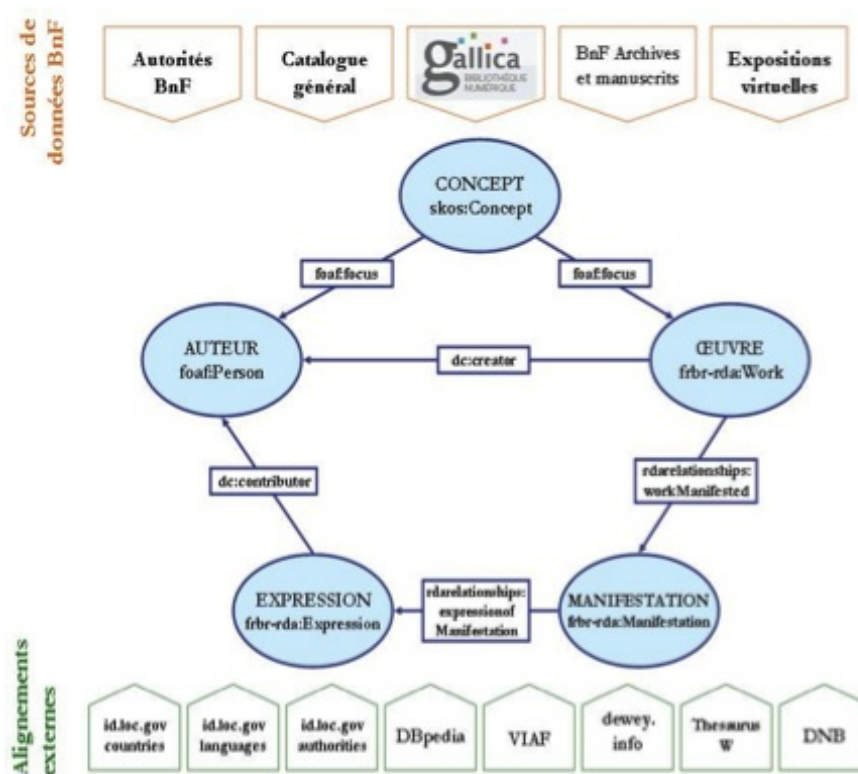


Figure 9: FRBR-based general model of data.bnf.fr
(Bibliothèque nationale de France, 2013)

3 Stage 3: Documenting Processes for Reuse

Some cultural heritage organizations may not finish a project, but leave behind excellent documentation for how to convert their catalog into linked open datasets and develop tools to encourage their use. Heres one organization doing just that:

3.1 The New York Times

3.1.1 Introduction

Continuing the legacy of the New York Times', which stretches back nearly to the founding of the newspaper, the New York Times and The New York Times Company Research & Development Lab have adopted linked open data to maintain and share the newspapers extensive holdings. The New York Times suite of LOD datasets, tools, and APIs are based in large part on the newspapers 150-year-old controlled vocabulary, which was released as 10,000 SKOS subject headings in January 2010.

The New York Times publicizes its projects through its blog, Open: All the News Thats Fit to print(f), and through social media. In addition to creating prototype tools such as Who Went Where, The New York Times also promotes the use of its APIs and source code of its tools. Open has been a regularly updated blog since the New York Times Company began its foray into the use and promotion of open source software in 2007.

3.1.2 The Dataset

The foundation of the New York Times dataset is the New York Times Index, which was published quarterly beginning in 1913 and continues to be published today, although with less frequency. These red-covered volumes contain a cross-referenced index of all of the names, articles, and items that appear in the newspaper. Along with creating an authoritative controlled vocabulary, the New York Times Index also helped to establish the New York Times as a trusted research resource for students, scholars, and librarians throughout the United States. As the New York Times continues to promote and develop its LOD assets it continues its legacy as information innovators.

The New York Times began publishing its vocabulary as linked open data in 2009; by 2010 the vocabulary had grown to include 10,000 subject headings. Currently (as of March 2013) the dataset includes the names of 4,978 people, 1,489 organizations, 1,910 locations, and 498 descriptors, or 10,467 tags in total. These tags are available as RDF documents and as HTML. Individual data records can be browsed alphabetically, downloaded in packages of SKOS files, or queried using APIs.

3.1.3 The APIs

As of the spring of 2013, the New York Times has released fifteen APIs. These APIs range from a Movie Reviews API that searches movie reviews by keyword and links to Critics Picks, to the TimesTags API, which matches queries to the New York Times controlled vocabulary. The documentation for the suite of APIs is hosted in the Developer section of the New York Times website (<http://developer.nytimes.com/docs/read/timestags-api>), which includes a glance view of the API as well as suggested uses for each API and a forum for users and developers.

All of the New York Times APIs are available in a JSON response format and a smaller subset is available as XML or serialized PHP. The HTTP method universally used by New York Times APIs is GET.

3.1.4 Who Went Where Tool

The New York Times, in a partnership with R&D labs, has developed a prototype tool to encourage developers to utilize its dataset. The tool, Who Went Where, is a search engine that enables users to search for recent Times coverage of the alumnae of a university or college. In addition to an introductory blogpost that explains the step-by-step process of creating your own linked data application with New York Times LOD, they have also made the applications source code available to the public.

Who Went Where is a JQuery application that queries DBpedias SPARQL endpoint. The high-level application control as described on the New York Times Open blog is as follows:

1. The application starts by initializing an auto-complete field with the names of all the colleges and universities in DBpedia.
2. When the user selects the name of an institution from the auto-complete field, the application queries DBpedia for the NYT identifiers of all the alumni of that institution. These identifiers are then used to query the New York Times Article Search API for the ten most recent articles about each alumnus. Then we use a little jQuery magic to display and format these articles (Sandhaus, 2010).

Who Went Where is not only a unique application that showcases the value of the New York Times and its APIs, but also a unique and elegant of an elegant and relatively straightforward application of these resources. The power of this tool is amplified by the documentation surrounding it, including the source code, which is available freely online.

3.1.5 Analysis

The New York Times has paired an extensive dataset with approachable documentation to create a powerful tool in the promotion of both the LOD movement and the newspapers resources. The Who Went Where search tool created using one of the New York Times APIs is not necessarily the most exciting application of LOD, but it does present an attainable and elegant framework for the use of APIs.

3.2 Deutsche National Bibliothek

3.2.1 Project Summary

The Deutsche National Bibliothek (DNB) began to publish authority data as linked open data in 2010, adding bibliographic data in 2012. The goal of the LOD program is to publish its library data on the Web in a way that makes its semantic relationships available to other data sources. The DNB describes linked data as the appropriate conceptual framework for achieving this goal.

In Germany, authority file data about entities such as persons, corporate bodies, and the like are collaboratively collected and maintained. These data, as well as the DNB's bibliographic data, are relevant to many libraries and other cultural heritage institutions. Conceptual links from these data to several other national and international data sources exist, but are not always part of the data representation that is exchanged with other institutions.

The audience for the DNB's linked open data includes users of library data (currently authority file data about people, corporate bodies, and subject headings) in particular and users of the Semantic Web in general.

Project leaders include Jan Hannemann, who is listed as owner on the W3C page (World Wide Web Consortium, 2010) and Julia Hauser. There is a way to leave comment for Ms. Hauser at the DNB Linked Data Service Web page. The DNB is a federal institution with legal capacity under public law. The annual funds provided from the budget of the Minister of State for Culture and the Media currently amount to roughly €45 million.

3.2.2 Technology

The bibliographic data were converted to machine-readable format using existing ontologies. The modeling of the bibliographic data was based on internal format specifications for DNB and ZDB bibliographic data.

- Bibliographic Ontology (Bibo)
- International Standard Bibliographic Description (ISBD)
- Dublin Core
- RDF Schema
- OWL
- FOAF

Examples of fields used for the DNB Linked Data service are available in a publication, *The Linked Data Service of the German National Library: Modelling [sic] of bibliographic data* (DNB, 2012). The GND Ontology specification for authority names is available on the website as well.

As shown in the screenshots below, interested users can access the data either through the existing search interface of the library's Web portal or by downloading the data in the form of a database dump. The current portal has a search interface and allows you to use RDF/XML representation links (e.g. on appropriate person-related pages). The right column contains links to download metadata records, including the RDF/XML file.



Figure 11: DNB Author Record

```
<rdf:RDF
  xmlns="http://d-nb.info/gnd/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:gnd="http://d-nb.info/standards/elementset/gnd#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  <gnd:DifferentiatedPerson rdf:about="http://d-nb.info/gnd/118514768">
    <gnd:variantNameEntityForThePerson rdf:parseType="Resource">
      <gnd:forename>Bertolt</gnd:forename>
      <gnd:surname>Brecht</gnd:surname>
    </gnd:variantNameEntityForThePerson>
    <gnd:variantNameForThePerson>H. Brecht, H. Bertolt</gnd:variantNameForThePerson>
    <gnd:variantNameEntityForThePerson rdf:parseType="Resource">
      <gnd:forename>...</gnd:forename>
      <gnd:surname>Brecht</gnd:surname>
    </gnd:variantNameEntityForThePerson>
    <gnd:variantNameEntityForThePerson rdf:parseType="Resource">
      <gnd:forename>B.</gnd:forename>
      <gnd:surname>Brecht</gnd:surname>
    </gnd:variantNameEntityForThePerson>
    <gnd:professionOrOccupation rdf:resource="http://d-nb.info/gnd/4053309-8"/>
    <gnd:variantNameForThePerson>Brecht, Eugen Berthold Friedrich</gnd:variantNameForThePerson>
```

Figure 12: DNB RDF File for Authority Name

Additional data access options are the SRU (Search / Retrieve via URL) and OAI-PMH (the Open Archives Initiative Protocol for Metadata Harvesting). The two options can be used for different purposes. SRU allows users to conduct a specific search without the need for a database of their own, while OAI permits continuous synchronization of large amounts of data.

SRU utilizes CQL (Common / Contextual Query Language) for data query. Full specifications are available from the Library of Congress (Library of Congress, 2007). An overview of the OAI-PMH interface is available at Sourceforge (DNB, 2009).

All search and retrieval options are detailed in the DNBs Note Regarding Access (DNB, 2012). The full dataset is also available for downloading at <https://portal.dnb.de/opac.htm>. A SPARQL interface is not available. The data file is covered by a Creative Commons Zero license.

3.2.3 Analysis

DNB's Linked Data service rates five stars on Tim Berners-Lees Linked Open Data star system. The dataset is online, machine-readable, and fully accessible in a number of formats, including SRU and OAI-PMH web interface and RDF download. Each relies on open standards and established technologies and formats and is ready to be used and linked to other works.

DNB falls in Stage Three of our life cycle, as it offers its dataset and an extensive set of documentation and specifications for download, with limited thoughts on what external parties can do with it. The interfaces offered are primarily a way to access the bibliographic and authority data, and not necessarily a new way to promote the datasets; otherwise, it might be included in Stage 4.

4 Stage 4: Developing User Interfaces

Several organizations and informal groups have made headway in developing new user interfaces that allow those interested in culture heritage to experience open and linked collections in new ways. Many of these projects are still at a proposal stage, but highlight the work that has yet to be done and the challenges that will have to be met in order to integrate linked open data into every cultural heritage users experience.

4.1 Pan-Canadian Documentary Heritage Network

As described in Section 1, Developing Datasets, the Pan-Canadian Documentary Heritage Networks Out of the Trenches LOD project created and published datasets expressed in RDF/XML format that integrated information provided by its collaborating institutions. The partners then decided to create a visualization rather than implementing a search application. Models included projects by Australian Tim Wray, including The Real Face of White Australia (<http://invisibleaustralians.org/faces/>)

For the visualization, the story of a single soldier, Mike Foxhead, was selected. Resources specifically about him include newspaper clippings, a photographic portrait, and a diary page from his battalion on the day he died; related resources include photographs from the battle where he died and information about the native tribe of which he was a member. The visualization is then presented in a 5-minute video (http://www.canadiana.ca/sites/pub.canadiana.ca/files/LOD-Demo-ENG_0.mp4) that explains the process and shows the range and variety of information available. As the report notes, this visualization provides only a sampling of what is possible. . . The potential for discovery is limited only by the imagination of the user of the metadata (PCDHN, 2012, pp. 12-13).

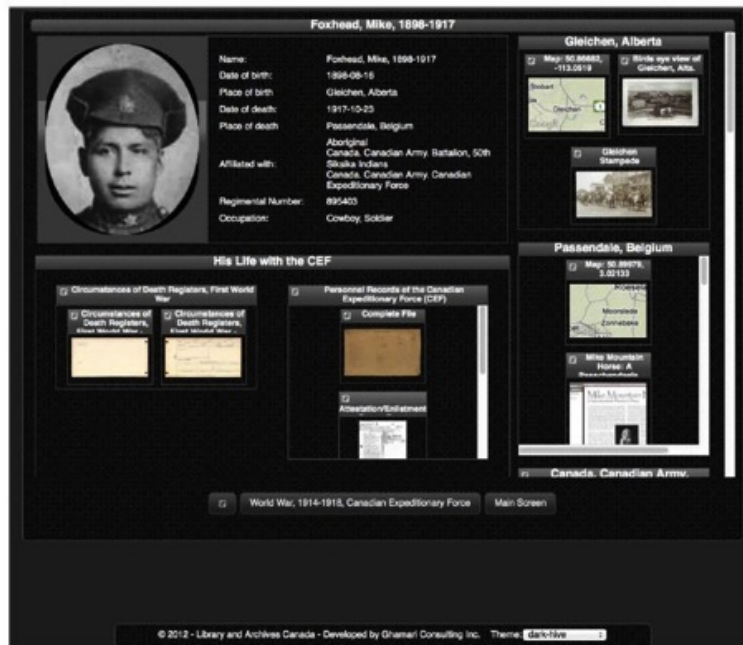


Figure 13: Screenshot from the Out of the Trenches demonstration interface showing the varied information from multiple sources related to the story of a single soldier (PCDHN, 2012, p. 11)



Figure 14: Screenshot from the Out of the Trenches demonstration interface showing the discovery dimensions available from the resources provided (PCDHN, 2012, p. 9)

The report finds that using RDF and LOD is a good approach for integrating resource discovery, allowing users to integrate PCDHN resources with those of other institutions to create their own stories and virtual exhibitions, effectively remov[ing] the constraints of existing web approaches in which the paths followed by web users are explicitly defined by the organization hosting the metadata/resources (PCDHN, 2012, p. 14). The partners also found that vocabularies can be repurposed, which extends the power and reach of the resources described using these vocabularies, giving as an example soldier as both a profession and as a subject heading, using a single URI rather than multiple URIs from different vocabularies. They believe RDF to be a powerful tool that makes integration across collections easy.

4.2 Agora Project

4.2.1 Introduction

The Agora project is a collaborative effort involving several Dutch cultural heritage institutions concerned with historical context and methods of manipulating and redefining context through social media platforms. One major aim of the project is to shift the viewpoint of historical narrative (or even the concept of history as a narrative) from that of the curator or institution to that of the viewer. Members of the project have expressed the desire to do away with the official version of history by creating apps that connect artifacts in disparate collections and allow users to link and discuss artifacts, locations and events as they see them. Partners for the Agora project include VU University Amsterdam, the Rijksmuseum Amsterdam, the Dutch national audiovisual archive Beeld en Geluid, and NWO Continuous Access to Cultural Heritage. Three teams comprise the project — the Science Team: VU History department, Susan Legene, and Chiel van der Akker; Computer Science Team: Guus Schreiber, Lora Aroyo, Marieke van Erp and Lourens van der Meij; and the Heritage Team: Rijksmuseum Amsterdam, Geertje Jacobs, Beeld en Geluid, and Johan Oomen (part-time PhD candidate). Eventing History is the tagline of the project and it plays on the concept of inventing history. The project aims to put the power of defining what constitutes a historical event into the hands of its apps' users. The current demo is available in a touchscreen or home computer version. The app includes a faceted search that allows users to see and make connections between artifacts in different settings and institutions. Unlike the immersion of curated multimedia presentations set in one institution the app allows users to define their own relationship to cultural heritage objects. Through linking related objects across institutions users can see objects in situ while this may no longer be possible in a physical space. The demo gives the example of objects belonging to a single individual spread across several institutions available for view together on one screen.

4.2.2 Authorities and Vocabularies

The Agora project uses B&G Open Images, 1,000 Videos, GTAA Thesaurus, 172,000 SKOS Concepts.

4.2.3 Analysis

Information on the Agora project is available through a Wordpress blog and is updated as papers and presentations are published. The project is ambitious in ideology and scope but technological documentation is not a strong point. Agora is as much concerned with the

philosophical constructs involved in cultural heritage description as it is with technological innovation. An example of this is evident in the current demo testing phase of the project. In the course of testing usability of the demo with groups of students it was discovered that a particular demographic, those of Dutch West Indies descent did not find the demo application useful for their research. This finding was of great concern to project and it has spawned papers and conference presentations while documentation on the development of the application itself has not increased at the same rate. Similarly, development of the project is partially fueled by the dissertation research of graduate students and the fate of the project beyond participant graduation is uncertain. Although Agora has begun some user interface development in the end Agora could make its mark on the LOD landscape more as philosophical or advocacy movement than as the producer of a usable application.

4.3 Amsterdam Mobile City App

4.3.1 Introduction

This project was proposed by a group of Dutch scholars who wanted to explore the potential for a smart phone application based on cultural heritage linked open datasets integrated with a phones native GPS and camera functions. (van Aart, Wielinga, & van Hage, 2010). Their proposal discusses the logistics of how such an app would work, requirements for front-end usability, and potential pitfalls in data retrieval. While other apps have been created using linked open data (and will be discussed later in this paper), this particular proposal gives an excellent overview of the possibilities of using a mobile interface to interact with linked and open cultural heritage data.

The Amsterdam Mobile City App is envisioned as a tool that allows visitors to the city to become their own self-directed tour guides. Each tourist can hold a phone up to a landmark site, use GPS to match that site to a geo-tagged item or items in the Amsterdam Museums collection, and pull up information about that cultural heritage artifact. The app would also allow the user to toggle back and forth between cultural heritage data provided by the museum and contextual LOD provided by DBpedia.

4.3.2 Datasets

The scholars behind the Amsterdam Mobile City App propose using the MultimediaN E-Culture dataset, which encompasses 200,000 objects from Amsterdam galleries, libraries, archives and museums expressed in more than 20 million triples. This dataset is not rich enough for end users on its own, so the writers propose linking it to DBpedia and several LOD ontologies, for example, the LinkedGeoData (LGD) initiative. In this way the data potential is rich, but the retrieval process needed to grab data from multiple sources, parse it, and display it on a users smartphone is cumbersome and slow.

4.3.3 User Interface



Figure 15: Mock-up of an augmented reality app envisioned by Chris van Aart, Bob Wielinga, and Willem Robert van Hage (2010)

The envisioned app would allow users to create self-guided walking tours of Amsterdam. A user could snap a photo of a historic site within the app, the app would recognize the site from visual and geographical data, and it would then query several linked databases to present a concise and interesting summary of that site matched with information about related cultural heritage artifacts.

In the image above, the user is looking at the real-life image of a famous building, paired with a painting of that building, pulled from the Amsterdam Museum's collection.

4.3.4 Moving Forward

This paper was written by scholars at VU University Amsterdam and Universiteit van Amsterdam, with funding from NoTube (now defunct), Agora, and the Poseidon Project. It will take future support from similar grantees to fund the actual creation of this app. The project might look to the INE and their Open Cultuur Data initiative to inspire a successful workflow.

4.3.5 Analysis

The scholars behind this imagined app explore one form of user interface that would make use of linked open data in a way that doesn't require the user to have knowledge of SPARQL or other back-end interfaces. They envision an app that is easy to use, easy to browse, and matches users' end needs and expectations. They do not offer a solution to the problem of load time, but warn that it will be a concern for future developers.

4.4 Australian War Memorial

4.4.1 Background

The Australian War Memorial (AWM) offers online access to its archives and collections through images and personal storytelling beginning with the World War I centenary. According to the museum website:

The Memorial’s purpose is to commemorate the sacrifice of those Australians who have died in war. Its mission is to assist Australians to remember, interpret and understand the Australian experience of war and its enduring impact on Australian society. The functions of the Memorial are detailed in section 5(1) of the Australian War Memorial Act of 1980 (Australian War Memorial, n.d.).

At a Canberra LODLAM minibar event in March 2012, Adam Bell, Web production manager for the AWM, outlined how the museum is working with Web scientists at the CSIRO, the Commonwealth Scientific and Industrial Research Organisation, Australia’s national science agency, and other science researchers as well as with the cultural sector, to develop a semantic tool to link data from its archives so that designers, researchers, and historians can use to help find new ways of building historical narratives. They are creating ontologies and hope to get librarians and archivists involved. Bell noted that Drupal has good taxonomy tools, which they use as a proxy to other services. Ultimately, using linked open data will encourage a wider population of people to explore the site, find information about their relatives, and participate through social networks.

4.4.2 The AWM and LOD

While the website does include many of the building blocks of linked open data, such as embedded RDFa vocabularies, we were not able to find downloadable vocabularies, APIs, or much openly available information about the project other than summaries of presentations at various conferences and informal LODLAM events. From these limited resources, we have been able to create a picture of what has been done to date and hints toward plans for further opening of the tools in the future.

What we can see at the public website is a search interface that allows simple and advanced search, including facets and related categories. Viewing the page code confirms that image objects have embedded schema.org and FOAF definitions. David Peterson, the project developer, refers to an API in his presentations discussing linked data, specifically around minute 48 of his Drupal Down Under 2012 presentation (Peterson, 2012); however it appears that the API is only being used internally at this time. Screenshots of these elements are shown below.



Figure 16: AWM Collection Object

```
<!DOCTYPE html>
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" version="XHTML+RDFa 1.0" dir="ltr"
  xmlns:content="http://purl.org/rss/1.0/modules/content/"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:ogp="http://ogp.me/ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:sioc="http://rdfs.org/sioc/ns#"
  xmlns:sioc_t="http://rdfs.org/sioc/types#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#" lang="en">
<head profile="http://www.w3.org/1999/xhtml/vocab">
```

Figure 17: AWM Namespace Declarations in HTML tag

```
<div class="row">
  <div class="span8" about="http://www.awm.gov.au/collection/P00620.006" vocab="http://schema.org/"
    typeof="ImageObject">
    <div class="figure">
      <span rel="contentURL">
        
      </span>
      <span property="name" content="AT SEA, C. 1916. STUDIO PORTRAIT OF A GROUP OF SERVICEMEN AND NURSES.
      LEFT TO RIGHT ARE: LIEUTENANT HAMILTON, LIEUTENANT NORMAN PELTON, LIEUTENANT WILLIAM JOHN CRAZE,
      SISTER GRAHAM, LIEUTENANT ...">
```

Figure 18: Example of Embedded FOAF Metadata from the AWM Project

The current team includes Robyn Van Dyk, AWM curator, Adam Bell, Web production manager, and Liz Holcolme, Web manager. David Peterson is no longer working on the project; however, a job posting for a Web developer skilled with Drupal, JQuery, and linked data was available as recently as October 2012. The AWM received funding through appropriations of the Parliament of Australia, which are reviewed and reappropriated annually. We were not able to locate a funding source that was specific to the linked data project and assume that funding comes directly from the Australian government.

4.4.3 Technology

AWM uses Drupal to create vocabularies. According to its developer (Peterson, 2012), there are a number of Drupal modules that allow you to perform Named Entity Extraction, including Open Calais, Yahoo! Term extraction, and Apache Stanbol, which is an open source version of Open Calais. These vocabularies are embedded into the HTML as schema.org tags. Examples of the microdata used include itemscope, itemtype, and itemprop (used for latitude and longitude coordinates).

In the presentation, Peterson describes an API that the museum is using to allow images from the collection databases to be automatically placed into the Drupal website with metadata tags. Peterson explains that each Web page is an API: using a php parser via JSON, the relevant RDFa data are pulled from the Web page and converted to GraphDB. Then the data get passed back as an array via TPL Dataflow. The live link that is created in the target site points to a page containing the same data as the original content on the museums site.

4.4.4 Analysis

The AWM site is a bit of an oddball in terms of true linked open data, since each page is separately coded and no RDF file appears to exist for the entire collection. This makes it difficult to state definitively where the site sits in Tim Berners-Lees LOD star system. The data are available on the Web as embedded structured tags within the HTML, using nonproprietary format and open standards; however, one would need to build an API to access the structured tags for use in another project, so there are limitations to its openness. Some resources, including a presentation that Peterson posted on Slideshare (Peterson, 2011), indicate that an API was in development at some point, but there is no reference to it on the AWM website and, as noted, Peterson is no longer working at the museum. From the presentation, it appears that the API was being used internally to connect artifact records to the website. So each page may actually rate four stars on Berners-Lees scale, but the collection as a whole may not fulfill the requirements of the first star.

In our life cycle model, the AWM would be considered an experimental Stage 4 project. There is definitely potential for this project to be developed into a fully realized LOD initiative. Since the memorial has continued to hire Web developers for this project, it will be one to watch.

5 Stage 5: Promoting a Culture of Reuse

Part of Tim Berners-Lee's original vision was that organizations that embraced linked open data would also embrace a culture of sharing and reuse. While one of these projects does not meet the technical requirements for an LOD project, both exemplify the spirit of collaboration and sharing essential to a successful endeavor.

5.1 Open Cultuur Data: Whats Possible with Open (Unlinked) Data

5.1.1 Introduction

In September 2011, the Dutch Heritage Innovators Network (INE; see Grob et al., 2011) began the Open Cultuur Data (<http://www.opencultuurdata.nl>) initiative to encourage cultural institutions to release their data under open standards and encourage users to develop new uses for this data. They facilitated the creation of datasets from eight organizations: the Rijksmuseum, Amsterdam Museum, EYE Film Institute Netherlands, National Archives, the Netherlands Institute for Sound and Vision, and National Heritage Sites of the Netherlands (Oomen, Baltussen, & van Erp, 2012).

The datasets were made public in time to be relevant to developers entering the Apps for the Netherland contest—a nationwide contest encouraging developers to create smartphone apps that would engage users with the rich heritage of the Netherlands. INE hosted several hackathons before the contest deadline, creating a supportive environment for developers to use the new open cultural heritage datasets in the creation of cultural heritage apps with a strong user interface. Thirteen apps were created during the initial contest, including three award-winners.

5.1.2 Dataset

Each institution released data under creative commons or similar open standards, but not all organizations released their data in RDF triples (see <http://www.opencultuurdata.nl/datasets/>). Since the initial project, additional institutions have added their data to the collection. Below are a few examples of the varying ways in which organizations released data about their collections.

The Amsterdam Museum released two datasets. One is comprised of 70,000 art objects from the Middle Ages to the present, reflecting the cultural history of the city; the metadata are available under Creative Commons as an Adlib API for developers that want XML data via the OAI-PMH protocol and as a Linked Open Dataset (Amsterdam Museum, n.d.). The other is a set of more than 5,500 items from the fashion collection, available in XML from

the same Adlib API. Adlib is a database software used by many museums around the world, and can be integrated with Drupal (Jurcello, 2011). The Amsterdam Museums dataset is particularly valuable, because like many museums, only 20 percent of its collection is on display at the museum, but the entire collection is recorded in the API.

The EYE Film Institute created a dataset of 170 Dutch films made from 1913 to 1940 with descriptive metadata, available as an ODS file. These films represent the birth of the film industry before World War II, including the films of star Annie Bos, and the highly acclaimed 1934 feature *De Jantjes*. The Institute points out that Dutch films were largely overshadowed by Hollywood after World War I, and the industry was shut down entirely by Germany during World War II. They release this dataset as a way to give new life to this largely forgotten part of the Netherlands cultural heritage.

The National Archive released several datasets, including a collection of 150,000 images taken by the Anefo Photo Agency of Dutch life between 1959 and 1989. The data are available via an OpenSearch API in XML format. The archives also created a user-friendly search engine for querying a collection of historic maps.

The Rijksmuseum released a set of more than 110,000 digital objects with descriptions via an OAI-PMH protocol and shared a Python script for users to query the data. The collection dates from antiquity through the nineteenth century, but excludes more recent works so that the collection can be shared freely without copyright restrictions.

While every institution made the datasets freely available to use, manipulate, and share, the datasets are not as intertwined as they could be because few organizations converted their catalogs to linked open data.

5.1.3 Award-winning Apps

During the Apps for the Netherlands hackathons in November 2011, thirteen apps were created using the Open Cultuur Datasets. Three were particularly noteworthy:

- **Rijksmonumenten.info** (<http://rijksmonumenten.info/>) This app allows users to browse more than 61,000 buildings in the Netherlands and take geotagged photos of each building to share via Wikimedia. It won an education award.
- **ConnectedCollection** (<http://www.opencultuurdata.nl/?p=583>) This app is targeted more toward the cultural heritage organizations themselves, and allows them to install a widget on their site that shows users related objects from partner institutions. They won funding to continue development.
- **Vistory** (<http://www.vistory.nl/what-is-vistory.shtml>); demo at <http://youtu.be/MXt4ExebHsA> This project used a linked open dataset of images and video. Users can discover historical films shot near their location, and contribute to the geotagging of

historic videos. Much like the Amsterdam Mobile City App proposed earlier, this app allows users to become their own self-guided historians.

5.1.4 Analysis

The Open Cultuur Data Initiative is a great example of the spirit of linked open data. Although not all the institutions met the technical specifications to be linked, they were incredibly open. The INE fostered a spirit of collaboration that led to finished projects that can be enjoyed by any tourist or resident of the Netherlands.

5.2 EUscreen

5.2.1 Background

By its nature, audiovisual media, particularly analog recordings, such as pre-digital television, radio, sound recordings and film, are difficult to access. Yet these media are essential components of European heritage, collective memory, and identity. EUscreen's Linked Open Data pilot was created to address the need to make these artifacts openly accessible to a wide audience of users, including students, scholars, and the general public.

EUscreen builds a network of content providers, standardization bodies, television research partners, and specific user groups to provide multilingual and multicultural access to television heritage. EUscreen is also conscious of rights issues and supports user-led demand and interest for services and content, including the development of use-case scenarios for different contexts (research, learning, leisure, and creative reuse).



Figure 19: Sample EUscreen Content

Content is organized in a number of ways allowing users to view individual items in the collection in fourteen categories, via virtual exhibitions curated by content providers such as EUscreen archives and broadcasters, or via comparative exhibitions that exploring themes

comparatively across Europe. Their content selection policy and metadata framework borrows from existing standards such as the metadata schema of the European Broadcasting Union to tag a multiplicity of content through Europe and encourage exploration, comparative study, and serendipitous discovery. As one of the main audiovisual content aggregators for Europeana, EUscreen and its collection is also connected to an online collection of millions of digitized items from European museums, libraries, and archives.

The EUscreen project consortium is made up of 28 partners and 9 associate partners from 20 European countries, led by the Netherlands Institute for Sound and Vision and the University of Utrecht. A Best Practice Network funded within the eContentplus program of the European Commission, EUscreen started as a three-year project in October 2009. FIAT/IFTA, the European Broadcasting Union, and the EDL Foundation provide additional funding. EUscreen is the follow-up project of Video Active, an online platform with 10,000 items about the history of European television.

Johan Oomen of the Netherlands Institute for Sound and Vision is the technical director for the EUscreen LOD Pilot project. Nikolaos Simou, Vassilis Tzouvaras, and Nasos Drosopoulos covered the coordination and core implementation of lod.euscreen.eu/resource/, with the support of the EUscreen project. Jean-Pierre Evain created the harvesting schema mapping to EBU Core ontology and Panos Georgantas worked on HTTP redirection. Contact information is available at the pilot website.

5.2.2 Technology

EUscreen uses a variety of technologies, including XML, RDF, EBU Core ontology, 4store triple store repository, and SPARQL query. The different metadata models of the contributing institutions are aggregated into a single EBU Core metadata structure and published to the EUscreen portal. From there Europeana aggregates the content and makes it available through its website.

EUscreen harvests metadata using its MINT Ingestion Platform, which is available to registered affiliated organizations. These providers map their metadata using MINT to a harvesting XML schema that was implemented based on EBUcore (an audiovisual metadata standard of the European Broadcasting Union). The standardized metadata allows for interoperability between the various partner collections. The EUscreen LOData Pilot website contains a link to a Google document showing the data mapping in spreadsheet format. This metadata is then expressed as EBU RDF triples stored in 4store triplestore, which can then be accessed by users (see diagram below).

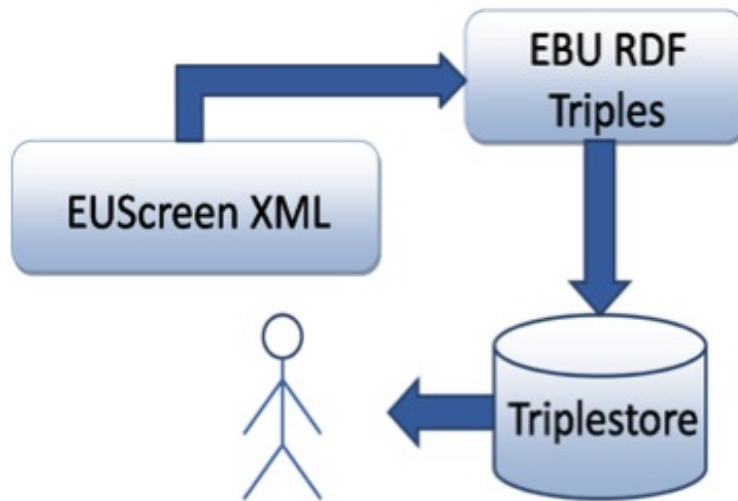


Figure 20: EUScreen metadata transformation
(<http://lod.euscreen.eu>)

A diagram illustrating an excerpt of the resources and the way that they are linked is shown below.



Figure 21: EUScreen RDF Schema

To provide access to the data, the lod.euscreen.eu/resource/ site serves data over the HTTP protocol, using established linked data recipes. HTTP URIs serve as identifiers and entry points into data. The data can also be consumed through the <http://lod.euscreen.eu/sparql> or the 4store repository interface via the Web. (We noted that these URLs were not working the last time we checked the website. The EUscreen Open Data Pilot website describes these resources as a pilot, so the Web page may change.)

All of this is seamless to the user of the EUscreen and Europeana websites or any site that aggregates EUscreen content. The EUscreen public-facing website has simple and advanced search as well as a graphical “Explore” interface that mimics advanced search functionality, and is somewhat easier to use than search (see screenshot below). Simple search box works as expected. Advanced search allows you to select facets to allow searching by genre, date, provider, publisher, color, geography, media type, and others.

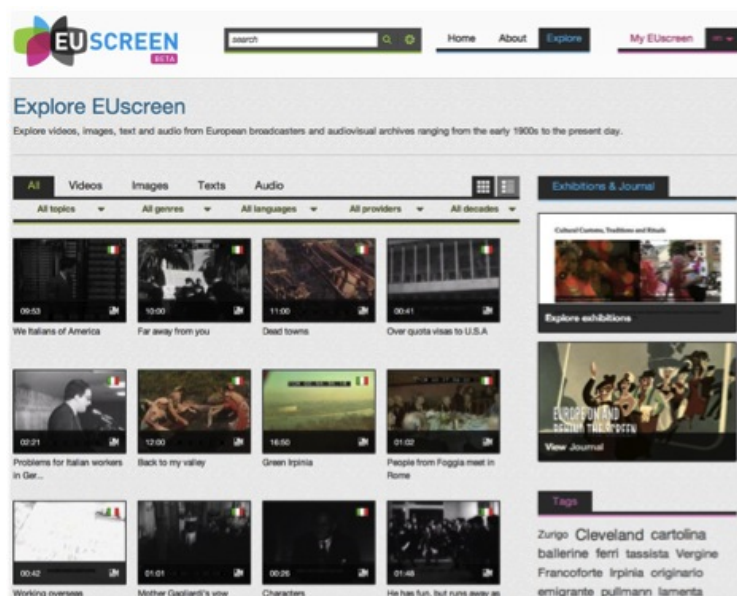


Figure 22: EUscreen Explore Interface
(<http://www.euscreen.eu/explore.html>)

5.2.3 Analysis

EUscreen rates five stars in the LOD star system. The dataset is machine-readable, non-proprietary and openly available under Creative Commons license. It uses RDF standards and is fully linked to external projects, notably serving as Europeana's main aggregator for audiovisual media.

In our life cycle model EUscreen is a Stage 5 project because it offers datasets and documentation, ideas for reuse, and lots of examples, all of which is actively promoted through linked data and cultural heritage conferences, exhibitions, and events. It seems natural that a project in the media and broadcasting field would be savvy in the promotion area. In addition, one of the project leaders is Netherlands Institute for Sound and Vision, which seems to have been involved in many LOD projects.

6 Stage 6: Expanding the Definition of Cultural Heritage

It can be argued that technological advances themselves are what shape our definition of cultural heritage. For example, the advent of the digital library in the late 1990s gave shape to collections that focused on hyperlocal moments in history (Dalbello, 2004). Linked open datasets have the potential to do just the opposite and expand the very nature of what cultural heritage can be, bridging the gap between online and offline collections.

6.1 Open Data for Resilience Initiative

6.1.1 Introduction

The Open Data for Resilience Initiative (OpenDRI) is an initiative by the Global Partnership Facility for Disaster Reduction (GFDRR) that seeks to build an open data sharing program. OpenDRI aims to reduce the impact of disasters by empowering decision makers with quality information and the tools to support their decisions. It is currently working on implementation in 25 countries worldwide to improve disaster and climate change resilience and is an excellent resource for geospatial information, data, and knowledge sources. The data can be used for many purposes: establishing baseline data, conducting risk assessments, planning, project monitoring, and tracking progress. The website is intended to facilitate more effective support to a countrys rehabilitation, recovery, and longer- term sustainable development.

6.1.2 Relationship to Cultural Heritage

Sharing data and creating open systems promotes transparency and accountability and ensures that a wide range of actors are able to participate in the challenge of building resilience. Not all countries have reliable census information or statistical data within their respective country or at the village level. In order to build resilient societies, policy makers and the public must have access to the right data and information to make informed decisions. Examples include decisions about where and how to build safer schools, how to insure farmers against drought, and how to protect coastal cities against future climate impacts. Citizens and governments can use OpenDRI in their planning process in face of disaster and protect the cultural institutions that make up the social fabric of their respective country.

6.1.3 Partnership and Funding

OpenDRI is a highly collaborative initiative. The GFDRR is a partnership of 41 countries and 8 international organizations. As a global partnership, OpenDRI is led and overseen by Robert Soden of the World Bank. The GFDRR Secretariat is led by Francis Ghesquiere, manager for the World Banks Disaster Risk Management Practice Group. Christoph Pusch is lead specialist for Disaster Risk Management and GFDRR deputy-manager, focusing on strategy, knowledge, and advisory services.

OpenDRI is financially supported by the World Bank and a multidonor trust fund, including contributions from the European Union (EU) and 17 donor countries. During the period 2006-2012, the project received \$350 million.

6.1.4 OpenDRI in Action: Data Reconstruction in Haiti

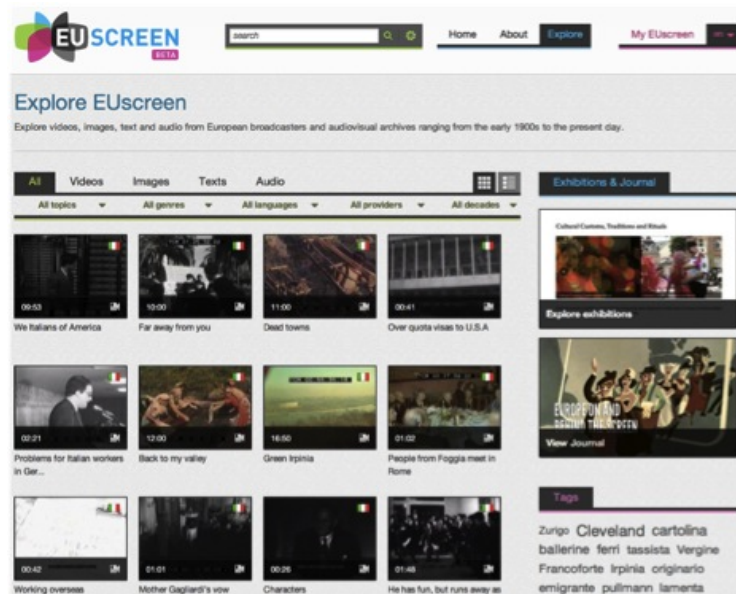


Figure 23: (Haiti Data, 2013)

The above example is a screenshot from Haitidata.org, which is a website created using Geonode, an open-source software tool. Following the 2010 Haiti earthquake, this risk assessment data was produced and made available for anyone to download and use. This tool produces realistic natural hazard impact scenarios for better planning, preparedness, and response activities. It provides a simple but detailed tool to combine data from scientists,

local governments, and communities to provide insights into the likely effects of future disaster events. The software is focused on examining the impacts that a series of hazard scenarios would have on specific sectors. This free open-source software tool enables individuals and organizations to contribute data and/or retrieve reliable data, thereby enabling people to collaborate in the recovery efforts for Haiti.

6.1.5 Analysis

By focusing on local high-risk communities, OpenDRI provides open access to information that will enable them to make more effective decisions on how to best minimize the impact of certain hazards on their livelihoods and cultural institutions. OpenDRI is a pilot with enormous potential to be innovative in addressing urban issues by creating data that is open and accessible to the public. It will be interesting to see how different stakeholders utilize this data in the coming years, and if OpenDRI's mission will actually provide better planning tools to help countries become more resilient and be able to recover quickly from disasters.

Conclusions

This survey examined the work of linked open data projects at many different stages of development. Rather than present one overarching conclusion, we present several ideas to consider:

1. Although the number of LOD datasets continues to increase, the actual use of LOD by cultural heritage institutions appears to remain limited at this time.
2. The nature of LOD pilot projects at cultural heritage institutions, while occasionally collaborative, seems to be highly curated and experimental. Many remain at the proof-of-concept stage, that is, an attempt by the institutions to see what is possible. Often, users cannot actually access the datasets or interfaces and documentation is limited, but the potential is there.
3. Trust remains an obstacle to the larger adoption of LOD. Computers can't tell, for example, that the statement "a pony is_a_kind_of fish" is untrue. It is likely that verifying the source of a statement when obtaining or presenting results will become a best practice, perhaps turning RDF triples into quintuples by adding a source label and value.
4. Published datasets hold great potential for making the content of an archive's collections known. For example, researchers studying Person A may discover that a collection of Person X's letters includes letters to or from Person A. Without linked data, this information might not have been easily discovered. Of course, this may require drilling down to item-level description, which is no longer typical in archival processing, so a challenge may be requiring a return to item-level description as a rule.
5. So far, no one project has embodied both the philosophical aims of LOD with the technological expertise. But more and more cultural heritage institutions are encouraging use of open data, even if they have not fully embraced the linking schemas. If the impact of the linked open data movement ends up being philosophical (with the language of RDF triples never widely adopted) is that still a success for the movement?

We posit that even with the limitations of linked open data projects — trustworthy metadata, limited funds, overburdened technical staff, and more proofs of concept than finished projects — each linked open data initiative pushes the potential of what cultural heritage can mean for users around the world. Each new interface and dataset makes it more possible for library, archive, and museum visitors to gain a richer experience and understanding of their surroundings.

In the for-profit world we see companies like Google adopting linked data to provide users with a richer experience. We maintain that the technical metadata and the philosophy of openness will continue to pervade the cultural heritage field and offer more opportunities for users to access cultural heritage objects outside the silos of museums, archives, and libraries.

References

- Agora. (n.d.). Retrieved from <http://agora.cs.vu.nl/>
- Amsterdam Museum (n.d.). *Open data*. Retrieved from <http://www.amsterdammuseum.nl/open-data>
- Archives Hub. (n.d.). Retrieved from <http://archiveshub.ac.uk>
- Archives Hub. (n.d.). *Linked Open Copac and Archives Hub (LOCAH)*. Retrieved from <http://archiveshub.ac.uk/locah/>
- Aroyo, L. (2012, December 13). *AGORA Creating the Historic Fabric for & Providing Web-enabled Access to Objects in Dynamic Historical Sequences* - ISAB 2012 Site Visit. Retrieved from <http://www.slideshare.net/laroyo/agora-project-final-review-2012>
- Archives Hub. (n.d.). *Linking Lives: Using linked data to created biographical resources*. Retrieved from <http://archiveshub.ac.uk/linkinglives/>
- Australian War Memorial. (n.d.) *About the Australian War Memorial*. Retrieved from <http://www.awm.gov.au/about/>
- Australian War Memorial. (n.d.) Search interface. Retrieved from <http://www.awm.gov.au/search/collections/?mode=advanced>
- Berners-Lee, T. (2006, last updated June 18, 2009). *Linked data*. Retrieved from <http://www.w3.org/DesignIssues/LinkedData.html>
- Bibliographic Ontology (Bibo). (n.d.). <http://purl.org/ontology/bibo/>
- Bibliothèque nationale de France. (n.d.). *Data.bnf.fr*. Retrieved from <http://data.bnf.fr>
- Bizer, C., Cyganiak, R., & Heath, T. (2007). *How to publish linked data on the Web*. Retrieved from <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/>
- Brickley, D., & Miller, L. (2010, August 9). *FOAF vocabulary specification 0.98*. Retrieved from <http://xmlns.com/foaf/spec/>
- Commonwealth Consolidated Acts. (1980). Australian War Memorial Act of 1980 – Sect 5. Retrieved from http://www.austlii.edu.au/au/legis/cth/consol_act/awma1980244/s5.html
- Culture Victoria (2012, April 24). *Linked open data – Melbourne workshop* [Web log post]. Retrieved from <http://blogs.cv.vic.gov.au/news/linked-open-data-melbourne-workshop/>
- Cyganiak, R., & Jentsch, A. (2007, last updated September 19, 2011). *The linking open data cloud diagram*. Retrieved from <http://lod-cloud.net/>
- Dalbello, M. (2004). Institutional shaping of cultural memory: Digital library as environment for textual transmission. *Library Quarterly*, 74, 265-298. Retrieved from <http://www.jstor.org/stable/10.1086/422774>
- The Data Hub. (nd). Hungarian National Library (NSZL) catalog retrieved from <http://datahub.io/dataset/hungarian-national-library-catalog>
- Deutsche National Bibliothek. (2009, last updated January 12). *OAI harvester manager web based user interface: Getting started*. Retrieved from <http://oaiharvestmangr.sourceforge.net/doc/guibasics.html>
- Deutsche National Bibliothek. (2012, April 26). *The linked data service of the German National Library: Modelling of bibliographic data*. Retrieved from http://www.dnb.de/SharedDocs/Downloads/EN/DNB/service/linkedDataModellierungTiteldaten.pdf;jsessionid=26D04FEA12CA977D65810052DF45D2BD.prod-worker2?__blob=publicationFile
- Deutsche National Bibliothek. (2012, April 30). *GNB ontology*. Retrieved from <http://d-nb.info/standards/elementset/gnd#>
- Deutsche National Bibliothek. (2012, June 6). *The linked data service of the German National Library: Note regarding access*. Retrieved from http://www.dnb.de/SharedDocs/Downloads/EN/DNB/service/linkedDataZugriff.pdf?__blob=publicationFile

- Deutsche National Bibliothek. (2013, February 13). OAI overview. Retrieved from http://www.dnb.de/EN/Service/DigitaleDienste/OAI/oai_node.html
- Deutsche National Bibliothek. (2013, February 13). SRU overview. Retrieved from http://www.dnb.de/EN/Service/DigitaleDienste/SRU/sru_node.html
- Deutsche National Bibliothek. (2013, last update April 11). *Linked Data Service of the German National Library*. Retrieved from <http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata.html>
- Digital Library Federation. (2011). *Civil War 150: Notes toward a linked data case study*. Retrieved from <http://www.diglib.org/forums/2011forum/schedule/civil-war-150-notes-toward-a-linked-data-case-study/>
- Dublin Core Metadata Initiative. (2012, June 14). *DCMI metadata terms*. Retrieved from <http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms>
- European Broadcasting Union (EBU). (n.d.). *Metadata specifications*. Retrieved from <http://tech.ebu.ch/MetadataSpecifications>
- The European Open Library. (nd). Hungarian National Shared Catalogue. Retrieved from <http://www.theeuropeanlibrary.org/tel4/contributor/P01210>
- EUScreen (n.d.). *About EUScreen*. Retrieved from <http://www.euscreen.eu/about.html>
- EUScreen blog. (n.d.). Retrieved from <http://blog.euscreen.eu>
- EUScreen harvesting XML schema. (n.d.). Retrieved from <http://lod.euscreen.eu/EUScreen.xsd>
- EUScreen ingestion tool. (n.d.). Retrieved from http://euscreen.image.ece.ntua.gr/euscreen/Login_input.action
- EUScreen. (n.d.). *EUScreen linked open data pilot*. Retrieved from <http://lod.euscreen.eu>
- Finin, T. (2009, October 30). *New York Times publishes linked open data* [Web log post]. UMBC ebiquity. Retrieved from <http://ebiquity.umbc.edu/blogger/2009/10/30/new-york-times-publishes-linked-open-data/>
- Freebase. (2010, last updated August 24). *CDW150*. Retrieved from <http://wiki.freebase.com/wiki/CWD150>
- Geonode. (2013). *Haiti data*. Retrieved from <http://haitidata.org/>
- Gheen, T. (2012, July 9). Library of Congress launches beta release of linked data classification [Web log post]. Retrieved from <http://blogs.loc.gov/law/2012/07/library-of-congress-launches-beta-release-of-linked-data-classification/>
- GlimwormIT. (2012, January 2). *Apps4NL: Vistory* [Video file]. Retrieved from <http://www.youtube.com/watch?v=MXt4ExebHsA&feature=youtu.be>
- Global Facility for Disaster Reduction and Recovery (GFDRR). (2013). *Open data for resilience initiative (OpenDRI)*. Retrieved from <https://www.gfdr.org/opendri>
- Grob, B., Baltussen, L., Heijmans, L., Kits, R., Lemmens, P., Schreurs, E., Timmermans, N., & van Tuijin, E. (2011). Why reinvent the wheel over and over again? How an offline platform stimulates online innovation. Paper presented at Museums and the Web 2011, Philadelphia, PA, April 6-9, 2011. Retrieved from http://www.museumsandtheweb.com/mw2011/papers/why_reinvent_the_wheel_over_and_over_again_how
- Horváth, A. (nd). Linked Data at the National Széchényi Library. Retrieved from http://nektar.oszk.hu/w/images/0/04/LinkedDataAtNszl_06.pdf
- Hungarian National Library. (n.d.). Retrieved from http://nektar2.oszk.hu/librivation_eng.html
- ISKO UK. (2010, May 15). Hungarian National Library OPAC and Digital Library Published as Linked Data. <http://iskouk.blogspot.com/2010/05/hungarian-national-library-opac-and.html>
- Jurcello. (2011, February 25). Adlibapi [Web log post]. Retrieved from <http://drupal.org/project/adlibapi>

- Kelley, M. (2011, October 31). Library of Congress provides details about transition away From MARC. *Library Journal*. Retrieved from http://www.libraryjournal.com/lj/home/892637-264/library_of_congress_provides_details.html.csp
- Library and Archives Canada (LAC). (2012, last updated July 11). *Towards a Pan-Canadian Documentary Heritage Network: Setting the course for information resources*. Retrieved from <http://www.bac-lac.gov.ca/eng/about-us/Pages/towards-a-pan-canadian-documentary-heritage-network.aspx>
- Library of Congress. (2007, July 25). Archive of SRU version 1.1 specifications. Retrieved from <http://www.loc.gov/standards/sru/sru1-1archive/>
- Library of Congress. (n.d.). *Library of Congress Linked Data Service: About*. Retrieved from <http://id.loc.gov/about/>
- Maarten. (2012, January 11). Demo: Peer-to-peer collective-ontsluiting met ConnectedCollection [Web log post]. Retrieved from <http://www.opencultuurdata.nl/?p=583>
- Mason, I. (2012, March 18). Canberra LODLAM minibar – Tuesday 27 March 2012 [Web log post]. Retrieved from <http://lodlam.net/2012/03/18/canberra-lodlam-minibar-tuesday-27-march-2012/>
- National Széchényi Library. (2011, April 22). Semantic Web. Retrieved from http://nektar.oszk.hu/wiki/Semantic_web
- Netherlands Institute for Sound and Vision. (n.d.) Retrieved from <http://www.beeldengeluid.nl/en>
- New York Times. (2009). *Linked Open Data*. Retrieved from <http://data.nytimes.com/>
- New York Times. (2013). *API documentation and tools*. Developer Network beta. Retrieved from http://developer.nytimes.com/docs/read/timestags_api
- New York Times. (2013). *Open* [Web log site]. Retrieved from <http://open.blogs.nytimes.com/>
- Oomen, J. Baltussen, L., & van Erp, M. (2012, April). Sharing cultural heritage the linked open data way: Why you should sign up. Paper presented at Museums and the Web 2012, San Diego, CA. Retrieved from http://www.museumsandtheweb.com/mw2012/papers/sharing_cultural_heritage_the_linked_open_data
- Open Cultuur Data Initiative. (n.d.). Datasets. Retrieved from <http://www.opencultuurdata.nl/datasets/>
- Open Metadata Registry. (n.d.). *International Standard Bibliographic Description (ISBD) elements*. Retrieved from <http://iflstandards.info/ns/isbd/elements/>
- Pan-Canadian Documentary Heritage Network (PCDHN). (2012, July 11). *Pan-Canadian Documentary Heritage Network (PCDHN) linked open data (LOD) visualization “proof-of-concept.” Out of the Trenches: Linked Open Data of the First World War. Final report*. http://www.canadiana.ca/sites/pub.canadiana.ca/files/PCDHN%20Proof-of-concept_Final-Report-ENG_0.pdf
- Parliament of Australia. (2011). *Australian War Memorial funding*. Retrieved from http://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/BudgetReview201112/WarMemorial
- Peterson, D. (2011, October 13). Bringing history alive: Telling stories with linked data and open source tools. Slide Presentation. Retrieved from <http://www.slideshare.net/davidseth/bringing-history-alive-telling-stories-with-linked-data-and-open-source-tools>
- Peterson, D. (2012, January 17). Telling stories with linked data: Australian War Memorial [Video file]. Presentation at Drupal Down Under 2012, Melbourne. Retrieved from http://www.youtube.com/watch?v=5q_FaLgJqO0
- Plessner, A. (2013, April 5). The *New York Times* has prototyped a link-centric search engine video file]. *Beet.TV*. Retrieved from <http://www.beet.tv/2013/03/timeslinked.html>
- Presto Centre. (n.d.). *EUScreen*. Retrieved from <https://prestocentre.org/library/projects/euscreen>

- Rijksmonumenten.info. (n.d.). *Rijksmonumenten.info: Alle rijksmonumenten van Nederland in een iPhone App, Layar en op Google Maps*. Retrieved from <http://rijksmonumenten.info/>
- Ruddock, B., & Stevenson, J. (2011). Creating linked open data for library and archive description. *Multimedia Information & Technology*, 37(4): 19-20.
- Sandhaus, E. (2010, March 30). Build your own NYT linked data application [Web log post]. *Open: All the News That's Fit to print(f)*. Retrieved from <http://open.blogs.nytimes.com/2010/03/30/build-your-own-nyt-linked-data-application/>
- Smith, L. (2013). Trenches to Triples and Linked Data. Presentation at the UKAD13 Conference, London, 7 March 2013. Retrieved from www.ukad.org/documents/2013forum/LianneSmith.ppt
- Stevenson, J. (2012). Linking lives: Creating an end-user interface using linked data. *Information Standards Quarterly*, 24: 14-23. Retrieved from <http://www.niso.org/publications/isq/2012/v24no2-3/stevenson/>
- Stevenson, J., & Stevenson, A. (2011, May 25). Lifting the lid on linked data: Linked data and the LOCAH project. Presentation at European Library Automation Group (ELAG) Conference, Prague, Czech Republic. Retrieved from <http://www.slideshare.net/fullscreen/JaneStevenson/lifting-the-lid-on-linked-data/>
- Tangible Pasts? Questioning Heritage Education. (nd). Retrieved from <http://www.tangiblepasts.eu/downloads/positionpaper.pdf>
- Tift, S., & Jones, A. (2001, November 17). Dusting off the search engine. The New York Times. Retrieved from <http://www.nytimes.com/2001/11/17/opinion/dusting-off-the-search-engine.html>
- Törnroos, J., Mäkelä, E., Lindquist, T., & Hyvönen, E. (2013). World War 1 as linked open data. Semantic Web Journal. Retrieved from <http://www.semantic-web-journal.net/system/files/swj458.pdf>
- Trenches to Triples. (n.d.). Retrieved from <http://trenchestotriples.wordpress.com/>
- van Aart, C., Wielinga, B., & van Hage, W. (2010). Mobile cultural heritage guide: Location-aware semantic search. Proceedings of Knowledge Engineering and Knowledge Management by the Masses (EKAW 2010) Lisbon, Portugal, Oct. 2010. Retrieved from http://www.few.vu.nl/~wrvhage/papers/Mobile_cultural_heritage_guide_ekaw2010.pdf
- Van der Akker, C. (2012, September 6). History as Dialogue On Online Narrativity. Retrieved from <http://www.slideshare.net/ChielvdAkker/history-as-dialogue-14314531>
- Verbruggen, C. (2012, January 7). Zin en onzin van DH-onderzoeksinfrastructuren in het federale België retrieved from <http://www.digitale-geschiedenis.nl/abstracts/zin-en-onzin-van-dh-onderzoeksinfrastructuren-het-federale-belgi%C3%AB%C3%AB>
- Vistory. (n.d.). *What is Vistory?* Retrieved from <http://www.vistory.nl/what-is-vistory.shtml>
- World Wide Web Consortium. (2004, February 10). *RDF vocabulary description language 1.0: RDF schema*. Retrieved from <http://www.w3.org/TR/rdf-schema>
- World Wide Web Consortium. (2004, last updated November 12, 2009). *OWL web ontology language reference*. Retrieved from <http://www.w3.org/TR/owl-ref/>
- World Wide Web Consortium. (2010, last modified October 15). *Use case linked data service of the German National Library*. Retrieved from http://www.w3.org/2005/Incubator/ldd/wiki/Use_Case_Linked_Data_Service_of_the_German_National_Library
- World Wide Web Consortium. (2011, last updated September 15). *Use case Civil War Data 150*. Retrieved from http://www.w3.org/2005/Incubator/ldd/wiki/Use_Case_Civil_War_Data_150

World Wide Web Consortium. (nd). *Hungarian National Library published its entire OPAC and Digital Library as Linked Data*. Retrieved from
<http://lists.w3.org/Archives/Public/public-lod/2010Apr/0155.html>