Case study: towards a linked digital collection of Latvian cultural heritage

Uldis Bojārs

National Library of Latvia,
Mūkusalas iela 3, Riga, LV-1423, Latvia
{uldis.bojars@lnb.lv}

Abstract. The paper describes a linked digital collection developed at the National Library of Latvia. It highlights the interlinked nature of digital objects and will help users discover new information by exploring links between the objects. The object linkage is enhanced by annotating contents of selected types of objects resulting in a network of links between annotated objects and entities mentioned in their annotations. The collection is aimed both at the general public and researchers, and will provide a Linked Data interface and publish open datasets that researchers may find useful. The paper also describes our experience in developing the system with the limited resources available.

Keywords: cultural heritage, digital collections, Linked Data, Rainis, Aspazija

1   Introduction

Some digital collections show each object separately and do not consider how these objects may be related to one another. We contrast this approach with a different kind of collections - Linked Digital Collections - that make extensive use of links between objects and create new ways for users to explore them.

This paper presents the "Rainis and Aspazija 150" digital collection that is being developed at the National Library of Latvia. It lets users explore various types of content and emphasizes the interlinked nature of the collection. Especially interesting in this regard are annotated transcripts of letters included in the collection that contain references to the objects referenced in the letters thus enriching the collection with new objects and relations.

As discussed in Section 3 the collection will publish information about objects and links between them as Linked Data. Throughout the paper we refer to the Linked Data principles introduced by Tim Berners-Lee [2]:

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.

2   "Rainis and Aspazija 150" digital collection

This paper describes a case study based on the "Rainis and Aspazija 150" Linked Digital Collection pilot project, currently in development at the National Library of Latvia (NLL).[1]

The collection is devoted to the 150th anniversary of famous Latvian poets and politicians Rainis and Aspazija (in 2015). Its goal is to collect a wide variety of types of digital objects and to present these two great personalities in a wider context. The pilot project provides the possibility for users to explore annotated works of both poets, a small part of their correspondence and various related materials (documents, photos, posters, audio and video files). The poets' mutual correspondence spanned 35 years, covering essential stages of Latvian history and culture development, and has been included in the Latvian national register of the UNESCO "Memory of the World" programme.

The pilot project is not aimed at quantity but rather at variety and quality (collecting different kinds of objects from multiple participating organizations and describing and interlinking these objects as accurately as possible). The current collection consists of 500 digital objects of various types: 84 literary works, 158 letters along with their annotated transcriptions, 120 photos and various archival documents, posters, audio and video recordings.
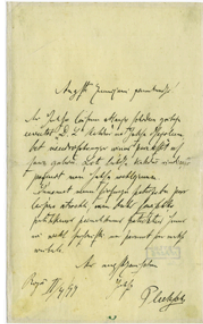
2.1   Annotations

The collection's objects may be linked to other related objects. Especially interesting in this regard are annotations that describe letters and some of the literary works in the "Rainis and Aspazija 150" collection. The letters that are a part of the collection had been previously transcribed and commented by experts helping us to disambiguate relevant entities mentioned in annotations and to link them to other information about these entities. Transcriptions of letters were annotated manually in order to ensure that the annotations are as precise as possible.

As a result of this annotation process new entities (such as people, organizations and events) and links to these entities were added to the collection. Where possible we tried to find existing digital objects that are related to or represent things mentioned in these annotations and to add these objects to the collection. For example, some letters mentioned a portrait of Aspazija to be published on the cover of a calendar. The expert annotating the letter found the portrait in a digitised version of the calendar and added an annotation linking to it.

Figure 1 shows the collection's page about a letter (from Rainis to Aspazija) linked to its annotated transcription. Figure 2 shows a fragment of an annotated letter that links to various entities mentioned in the letter: a concept for Zeus (greek deity), locations (Rainis' new and old work cabinet) and a person

---

[1] At the time of writing this collection was work-in-progress. It will be available at `http://runa.lnb.lv` from June 2016.

Raiņa vēstule Aspazijai Rīgā, 1894. gada 4.februārī

Digitālā objekta ID	60312
Resursa virstips	Teksts
Valoda	latviešu
URI	http://dom.lndb.lv/data/obj/60312
Autors	Rainis, Jānis, 1865-1929
Nosaukums	Raiņa vēstule Aspazijai Rīgā, 1894. gada 4.februārī
Oriģināla radīšanas datums	1894-02-04
Anotācija	Vēstule. Pliekšāns (Rainis) - Aspazijai. Rīga, 1894. gada 4.februāris. 1 lpp. rokr. ar melnu tinti. Papīrs dzeltējis.
Priekšmets	Rainis, Jānis, 1865-1929, Sarakste; Aspazija, 1865-1943, Sarakste; Vēstules
Saite no	Raiņa vēstule Aspazijai Rīgā, 1894. gada 4.februārī : teksts un komentāri - Rainis, Jānis, 1865-1929
Organizācija, kurā atrodas oriģināls	Rakstniecības un mūzikas muzejs
Objekta saite	http://runa.lnb.lv/60312

Fig. 1. A scanned letter from the "Rainis and Aspazija 150" collection.

vairāk žēl nekā šīs zaudētās dzīves. Manās rokās, kuras Tevi apskāva, vēl palikusi salda jūta pāri, it kā smarža paliek ilgu laiku, es visu aizvakaru vēl staigāju ietīts Tavā atmosfērā kā Dzevs[2] zelta mākonī. Man pa prātu plivinājās Tavi vaļējie mati, kas bija izkaisījušies pa visu spilvenu, es viņus salasu kopā pa vienam matiņam, un tam laikam prātam nav cita darba. Ja es arī esmu paša ievadraksta dziļumos vai politikā, - pie katra vārda man iegādājas kāda jauka atmiņa, kura tur it nemaz un it nemaz nav vietā. Es nevaru rakstīt par jauno kabinetu Francijā , man jāiedomājas mūsu laimīgā kabineta,[4] un tad, zināms, nāk atmiņā arī visas citas labās, klusās istabas. Iz rindiņām, kur prātojums par Bismarku[5],

Fig. 2. Annotated transcription of a letter (fragment) with links to related objects.

(Otto von Bismarck). The collection also has webpages describing these entities, including links back to other objects that reference (e.g. mention) them.

The entities mentioned in annotations are associated with URIs representing other resources related to these entities. One of the challenges in developing this system was how to maintain information about entities mentioned in annotations. Due to the limitations described in later sections we chose a simple approach where annotations are stored as links in HTML files and the information about entities is maintained in a table containing object type and related URIs. We hope to replace it with a better system for maintaining information about the entities of interest such as BBC Things - a Linked Data system that maintains information about all the things BBC systems may need to refer to.[2]

---

[2] http://www.bbc.co.uk/things/about

2.2   Use Cases

This section outlines some potential use cases for the collection. Regular visitors (such as the general public) can use the collection just like any other collection with the added benefit that additional links between objects would make it possible to "travel" through the collection and discover new content. The collection also provides visual tools - a timeline of the poets' works and a visualization of links between objects. The experience of Centre Pompidou shows that users may perceive a collection that consists of various types of objects interlinked to one another as a completely new way of exploring data: "... users who were just browsing the site out of curiosity liked the fact that they would get lost and discover unexpected resources" [1].

Advanced visitors (researchers, students) may engage in a more targeted exploration, looking at particular kinds of objects and links. They may also be interested in using the collection as raw data for further analysis (e.g. network analysis of the link graph or the content analysis of digital objects). These researchers might make use of all kinds of data and interfaces that the collection provides (such as a Linked Data interface, open data dumps, queries over the collection's data, and visualizations).

During project demonstrations users were especially interested in visualizations that show a network of links around the selected object. Regular users viewed it as a novel way for exploring the collection while researchers were quick to focus on some links in particular and to provide a larger context behind these links (e.g. that a literary work has a link to a person because she was an inspiration for the main character in this work).

Computer applications (and their developers) may access the collection in order to present or analyze its information in novel ways. They may use all the datasets and programming interfaces that the collection provides (Linked Data interface, APIs, open datasets). Open data catalogs may collect information about available open datasets. In order to do so they will need access to machine-readable information about open datasets provided by the collection which can be published using the Data Catalog (DCAT) vocabulary [3].

Types of data that could be published by a digital collection such as "Rainis and Aspazija 150" include the collection's object metadata including links between objects (e.g. published as Linked Data), an object link graph (for use in network analysis tools) and the contents of the collection's objects.

3   Describing the Collection as Linked Data

The collection will be published as Linked Data. The main question here is not how to publish the data but how to represent the information contained in the collection as RDF. There are two main approaches: (a) developing your own RDF vocabulary; or (b) reusing an existing RDF vocabulary or vocabularies.

Both approaches are valid – the Swedish Union Catalogue (LIBRIS) reuses existing vocabularies while datos.bne.es and Centre Pompidou have developed

their own vocabularies [4, 5]. The benefits of developing your own vocabulary are that it can more precisely and fully reflect the information that the system contains (especially important if the system stores data as RDF). The benefits of reusing existing vocabularies are that other applications might make direct use of the information published by the system.

We decided to reuse existing vocabularies where possible and to express the collection in RDF in two different ways: (1) using a combination of BIBO, FOAF and Dublin Core similarly as done in LIBRIS; and (2) using the Schema.org vocabulary. This two-pronged approach was chosen because BIBO, FOAF and Dublin Core offer more precise properties for expressing our LDC's information while Schema.org potentially offers more reach because it is used by major web search engines.

```
<http://example.org/letter/letter_01_URI>
    a bibo:Letter ;
    dc:creator <http://example.org/person/Rainis> ;
    bibo:recipient <http://example.org/person/Aspazija> ;
    bibo:place <http://example.org/location/Riga> ;
    dc:description "a letter from Rainis to Aspazija. Riga, 1894.02.04"@en ;
    dc:date "1894-02-04"^^xsd:date .

# Links to entities mentioned in the letter
<http://example.org/letter/letter_01_URI>
    dc:references <http://example.org/person/Aspazija> ,
                  <http://example.org/work/LNC04-000331919> .
```

The code listing shows an RDF data example describing a letter and its links to the entities mentioned in this latter. The example uses BIBO, FOAF and Dublin Core vocabularies and Turtle RDF serialization. Data about the entities mentioned in annotations contain links to other relevant Linked Data sources such as VIAF making the collection a part of a larger web of Linked Data.

4 Development: Constraints of the Pilot Project

What sets this case study apart from more research-oriented projects is that it had to deal with serious "real-life" constraints – the project was developed using library's internal resources (except for the data collection activity which involved other partner institutions) and did not have a separate budget or research partners that would help develop a custom state-of-the-art solution.

While these constraints seriously limited what could be developed in the pilot project and what technologies could be used, they also gave insights into the obstacles to the development of novel application with limited resources and into how developers not familiar with the Semantic Web perceive Linked Data principles and technologies. We hope that these observations will be useful to other organizations that may be in a similar situation.

The developers, not being familiar with the Semantic Web, prefer the technologies that they understand and are comfortable with (e.g. MS SQL, .Net and

C#). On the other hand, they know the library's existing systems and are good at integrating the pilot project with these systems. The system was developed on a relational database rather than RDF store. It was an acceptable choice because the data we are working with have a well-defined structure and come from existing library systems. The Linked Data interface is being built on top of this system.

We observed that developers found some Linked Data concepts easier to understand than others: that it is important to link objects together, to reuse the information available on other systems and, to some extent, the importance of URIs as identifiers for things. They might not consider the RDF data model an integral part of the system and view Linked Data as just one of the interfaces through which the system may publish its data.

A key challenge to the adoption of Semantic Web technologies in cultural heritage organizations that usually do not have sufficient research and development capacity is how can developers start using Semantic Web technologies when they spend most of their time on the existing legacy systems and technologies.

5   Conclusion

The "Rainis and Aspazija 150" digital collection introduced in this paper brings together various types of digital objects related to the topic of the collection and links these objects to one another. Of particular interest are the letters contained in the collection, transcribed and annotated with links to entities mentioned there, thus creating a new layer of links between collection's objects. The information about collection's objects and links between them will be made available as Linked Data.

The collection is about well-known and active Latvian cultural personalities and politicians and its contents will be interesting both to the general public and to researchers who previously did not have such a collection readily available.

References

1. Bermès, E.: Following the User's Flow in the Digital Pompidou. Linked Data and User Interaction, 162, 19 (2015)
2. Berners-Lee, T.: Linked Data - Design Issues. W3C (2006). Retrieved March 4, 2016, `https://www.w3.org/DesignIssues/LinkedData.html`
3. Maali, F., Erickson, J., Archer, P.: Data Catalog Vocabulary (DCAT). W3C Recommendation (2014), `https://www.w3.org/TR/vocab-dcat/`
4. Malmsten, M.: Making a library catalogue part of the Semantic Web. Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications. Dublin Core Metadata Initiative (2008)
5. Vila-Suero, D., Villazón-Terrazas, B., Gómez-Pérez, A.: datos.bne.es: A library linked dataset. Semantic Web, 4(3), 307-313 (2013)