NETWORK DIGITAL HERITAGE

A distributed network of heritage information

White paper

May 2017

White paper: A distributed network of heritage information - May 2017

The Dutch Cultural Heritage field is working together in a joint effort to improve the usability of their cultural heritage collection data. Our challenge is the development of a digital heritage infrastructure by avoiding aggregating and postprocessing of the data. Instead we aim to realize a true distributed network of digital heritage information. This paper will focus on our approach of developing a new, cross-domain discovery infrastructure for the Dutch heritage collections. With this new infrastructure we aim to improve the usability of the collection data at the source by implementing Linked Data principles in the collection registration systems. We urge the maintainers of the collections to align their data with formal Linked Data resources, like thesauri (people, place, periodes, concepts) and to publish data as Linked Open Data.

The Dutch Digital Heritage Network (NDE) program is a national program aimed at increasing the social value of the collections maintained by the Libraries, Archives and Museums in the Netherlands. The partners in the NDE network are the Ministry of Culture, the National Library, the National Archives, the Sound and Vision Institute, the Cultural Heritage Agency and a number of Research Institutes for Dutch Culture and History. These parties will formalize their co-operation through the installment of a new organization that will responsible for the realizing a joint strategy program for the Dutch cultural network. The goal is a distributed network build on the institutes and their stakeholders (including commercial parties), each contributing from their own perspective. The program is set up based on a three-layer model with a functional division between the management of data collections ('sustainability'), facilities for connecting that data ('usability'), and applications for presentation and use ('visibility').

Our work at the usability level is focused on the development of a lightweight cross-domain infrastructure that is built on a distributed architecture when possible. The core functionality consists of a network of terms that references all common definitions for places, people, concepts, time periods. These terms are made accessible through an OpenSKOS search API that collection registration systems can implement in order to search for relevant terms when annotating their cultural heritage objects. As a result the URIs of the terms will be added to the object descriptions. The NDE program works on getting all relevant thesauri available as Linked Data and provide facilities for term alignment and even support building new thesauri. Several tools for this work (CultuurLink[1] , PoolParty[2]) are being provided by the NDE network.

Having cultural heritage institutions publish their data as Linked Open Data with alignments to central definitions for places, persons, time periods and concepts is one part of the challenge. The other part is to provide means for browsing in a cross-domain, user centric fashion. Based on possible relevant URIs identified in the user query we would like to be able to browse the available linked data in the cultural heritage network. In general the concept of 'browsable linked data' is still a challenging concept. Although Tim Berners-Lee describes the concept of Browsable Graphs and states even that statements which relate things in two documents must be repeated in each, this is not a common practice in the Linked Data world[3]. If browsable Linked Data is offered then it is limited to the 'follow your nose' principle which is only based on using *forward links*. In order to really navigate in a bidirectional way through the LOD cloud, support for navigating using *back links* is needed as well. To our knowledge little research has been done so far on in this area.[4] [5]

Most Linked Data projects make bidirectional navigation work by aggregating Linked Data dumps and load them in a triplestore where both sides of the triples can be queried. Since our quest is

developing a distributed network that avoids replication of data and building very large central infrastructures this direction is undesirable.
An alternative approach is supporting federated queries. In general this would mean that every node in the network should support a SPARQL endpoint, which is a big challenge for small organizations and leads to major performance issues. Even a lightweight solution as Linked Data Fragments, developed by Ruben Verborgh and colleagues[6], which we hope to implement in many institutions, still leaves us at the question which endpoints have relevant data for a specific user question. The Dutch Digital Heritage Network consists of about 1500 institutions that hold collections. Random querying all the endpoints in this network using Linked Data Fragments would be impractical and unrealistic.

Therefore, we decided to follow a different route and build a (preferably distributed) registry that records the back links for all the terms used in the Digital Heritage Network. The registry contains the formal Linked Data definition of all the organizations and a high level description of the datasets they are maintaining, similar to the general CKAN registries for Linked Data (like datahub.io). In addition to this, we will also record *object profiles* that describe the relations between the object in the collection and the term URIs used in the object description. This information will provide the back links and makes it possible to navigate from a term URI to the objects that have a relation with this term.

For the implementing the synchronization of these *object profiles* to the central registry we are investigating the work of Herbert Van de Sompel, Sarven Capadisli and others on protocols like Resource Sync[7], Linked Data Notification[8] and Webmention.[9] With this new approach we hope to move away from a traditional *repository centric* approach to a more *web centric* approach where optimizing the usability of resources in their original environment is the starting point. We are currently developing a Proof-of-Concept for the distributed network of Dutch Heritage information and we will be able to show the first results in December 2017 at the SWIB conference.

---

[1] http://cultuurlink.beeldengeluid.nl
[2] https://www.poolparty.biz/
[3] https://www.w3.org/DesignIssues/LinkedData.html
[4] M. Stefanidakis et al., *Linking the (un)linked data through backlinks* (2012)
[5] M. Salvadores et al. *Domain-Specific Backlinking Services in the Web of Data* (2010)
[6] http://linkeddatafragments.org/
[7] http://www.openarchives.org/rs/1.1/resourcesync
[8] https://www.w3.org/TR/ldn/
[9] https://www.w3.org/TR/webmention/