DIVEplus: Browser for Integrated Linked Media

Oana Inel[2], Lora Aroyo[2], Jaap Blom[1], Victor de Boer[1,2], Annemarie Collijn[2], Werner Helmich[4], Christiaan Meijer[5], Johan Oomen[1], and Elco van Staveren[3]

[1]Netherlands Institute for Sound and Vision, Hilversum, the Netherlands
joomen@beeldengeluid.nl, jblom@beeldengeluid.nl

[2]Vrije Universiteit, Amsterdam, the Netherlands
v.de.boer@vu.nl, lora.aroyo@vu.nl, oana.inel@vu.nl, a.collijn@gmail.com

[3]Dutch National Library, the Hague, the Netherlands
elco.vanStaveren@kb.nl

[4]Frontwise, Utrecht, the Netherlands
werner@frontwise.com, dennis@frontwise.com

[5]The Netherlands eScience Center, Amsterdam, the Netherlands
c.meijer@esciencecenter.nl

ABSTRACT

DIVE is a linked-data digital collection browser, developed for providing an integrated, innovative and interactive access to objects from various heterogeneous online collections. The DIVE demonstrator uses events and event narratives as context for searching, browsing and presenting of individual and group of objects. DIVEplus extends the innovative DIVE approach in four ways: (1) add two new collections (i.e., Amsterdam Museum and Tropenmuseum), (2) integrating links to external linked open datasets (i.e., DBpedia, AAT and ULAN), (3) designing an intuitive way to deal with event narratives; and (4) automating the crowdsourcing of event annotations of the multimedia collection objects. The innovative interface combines Web technology and theory of interpretation to allow for browsing this network of data in an intuitive "infinite" fashion. The overall goal of DIVEplus is to gain insights on the scalability, robustness and reusability of the DIVE digital hermeneutics approach by testing its usability in user studies with a variety of end users (both research scholars and general audiences).

Keywords

Heterogeneous Linked Data, Crowdsourcing, Digital Hermeneutics, Historical Events, Digital History

1. DIVEPLUS DEMONSTRATOR

This demo presents DIVEplus[1], a tool which advances the way in which researchers and general audience interact with online heritage collections by allowing an integrated exploration of objects of heterogeneous collections. As such, DIVEplus extends the digital hermeneutics approach [2] of DIVE applied for cultural heritage collections and uses historical events and event narratives as context for searching, browsing and presenting collection objects [1]. It builds on the DIVE demonstrator[2], where semantics from existing heritage collections and linked data vocabularies are used to link objects with events, people, locations, times and other concepts that are depicted or associated with those objects. The main focus in DIVEplus is to provide support to both digital humanities scholars and general audience with an interest in history in their online explorations.

2. DATA SOURCES AND ENRICHMENT

The DIVEplus demonstrator allows for browsing of heterogeneous linked datasets as long as they contain media objects (images or videos) which are enriched with entities such as events, persons, places and other concepts. Currently, content from *three cultural heritage institutions* are enriched, linked and made available. The collections were selected from roughly the same period and topics to ensure that links between the collections could be established.

- *Dutch news broadcasts* form the Netherlands Institute for Sound and Vision (NISV)[3]. Within the DIVEplus we ingested a randomly selected subset of about 300 videos from the NISV collection of broadcast videos published as Open Data on the Openimages platform[4] from the period 1920-1980.
- *ANP Radio News Bulletins*[5] from the Dutch National Library (KB)[6]. In the DIVEplus we ingested 2210 KB digitized typoscripts (radio news scripts, to be read during news broadcasts) from the period 1937-1984.
- *Cultural heritage objects* from the Amsterdam Museum (AM)[7]. In DIVEplus we ingested 3500 images representative of the period 1950-1980.

Additionally, in the DIVEplus triple store we extended the existing cultural heritage linked data cloud with an automatic alignment of the enriched metadata from the above collections with various structured vocabularies, e.g. Gemeenschappelijke Thesaurus Audiovisuele Archieven (GTAA) and Amsterdam Museum Thesaurus, Persons list and Geo vocabulary. Thus, the collections made available are interlinked in a common linked data network of events, persons, places and concepts, which provides context for browsing and exploration of the cultural heritage objects.

---

[1]http://diveplus.beeldengeluid.nl
[2]http://dive.beeldengeluid.nl

[3]http://www.beeldengeluid.nl
[4]http://openimages.eu
[5]http://radiobulletins.delpher.nl/
[6]http://www.kb.nl
[7]https://www.amsterdammuseum.nl

Figure 1: DIVEplus Screenshot: shows a person (Berlage), with two related events and KB media objects associated with them (transcripts from two ANP radio programs), related location (Bilt) and related person (Dutch Queen). In the bottom, objects are also projected on a timeline (if timestamp is available)

3. RESULTS

The textual descriptions and descriptive metadata for all three collections are retrieved and converted to RDF. Our methodology incorporates a hybrid workflow for event enrichment in online collections. In this hybrid workflow the machines and the crowd collaborate in the process of extracting relevant events and event-related concepts. The results from the different tools and crowdsourcing, performed through the CrowdTruth platform[8] are consolidated to RDF. The data is modeled using the Simple Event Model (SEM) [3]. This model allows for the representation of events, actors, locations and temporal descriptions. We extend SEM with other Linked Data schemas, e.g. DC, SKOS, OpenAnnotation and FOAF to represent other types of resources linked to the media objects. Links are also established to external sources, including Wikipedia and DBpedia. The resulting dataset is stored in an RDF Triple store, which provides a SPARQL endpoint[9].

In the current triple store we host over 2 Million triples for the 5,000 Media Objects. These are annotated with 17,209 annotations to places, 5,044 actor-annotations and 2,992 separate event-annotations. 8,419 corresponce-triples (`skos:exactMatch`) are established between these entities, tying together the different vocabularies.

4. USER INTERFACE

DIVEplus implements an intuitive event-centric browsing interface for browsing online heritage collections by means of underlying linked data graph. The DIVEplus interface identifies as an engaging user experience that invites the users to continue the exploration at different levels of detail. Users become explorers by diving deeper into the data, as a diver deeper and deeper into an ocean trench discovering new species. This metaphor makes the interface a *"digital submarine"*. The design of the interface forms an innovative *"infinite exploration path"*, which unlocks the potential of touchbased explorative user interface. The DIVEplus *"infinity browsing"* interface is a combination of two core inter-

action concepts that involve a *horizontal level* supporting the breadth and a *vertical level* supporting the depth of information exploration and interpretation.

The interface acquires data from the data layer using the triple store's SPARQL API. Several queries are used to search entities by keyword, get related entities and get entity details. A smart image cache has been implemented to provide a visual representation for the relevant entities. Based on keywords from entity titles, images are retrieved from the five most relevant Wikipedia searches using the Wikipedia API[10] or the OpenCultuurData API[11] which covers an extensive set of Dutch open heritage- and cultural data. These images increase the user experience by supporting the visual navigation through the interface and the recognition of individual entities. Figure 1 shows the current version of the interface, optimized for tablets and modern web browsers.

We performed user studies with digital humanities professionals and students to test the usability of the DIVEplus approach. For future, we plan to perform more user studies within CLARIAH[12], to gain insights on the scalability, robustness and reusability of the DIVEplus digital hermeneutics approach, as well as its usability for history researchers and general audience.

5. REFERENCES

[1] V. de Boer et al. Dive into the event-based browsing of linked historical media. *Web Semantics: Science, Services and Agents on the World Wide Web*, 35(3):152–158, 2015.

[2] C. van den Akker et al. Digital hermeneutics: Agora and the online understanding of cultural heritage. In *Proc. of the 3rd International Web Science Conference. ACM, New York, NY, USA, 7 pages.*, 2011.

[3] W. R. van Hage et al. Design and use of the simple event model (sem). *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(2), 2011.

---

[8] http://crowdtruth.org/

[9] ClioPatria triple store data.dive.beeldengeluid.nl

---

[10] http://www.mediawiki.org/wiki/API:Main_page

[11] http://www.opencultuurdata.nl/api/

[12] http://www.clariah.nl/en/