DIVE+: Explorative Search in Integrated Linked Media

Oana Inel[2], Lora Aroyo[2], Jaap Blom[1], Victor de Boer[1,2], Werner Helmich[4], Carlos Martinez-Ortiz[5],
Liliana Melgar Estrada[3], and Johan Oomen[1]

[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

[3]University Of Amsterdam, the Netherlands
l.m.melgar@uva.nl

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

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

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

1. DIVE+ DEMONSTRATOR

DIVE+[1] 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 extends the digital hermeneutics approach [3] and uses events and event narratives as context for searching, browsing and presenting individual and groups of objects [1]. The innovative DIVE+ approach is many-fold: *(1)* integration of four heterogeneous cultural heritage collections (*i.e.*, news broadcasts from the Netherlands Institute for Sound and Vision, radio bulletins from the Dutch National Library, cultural heritage objects from Amsterdam Museum and Tropenmuseum), *(2)* integration of links to external linked open datasets (*i.e.*, DBpedia, AAT and ULAN), *(3)* intuitive way to deal with event narratives; and *(4)* automated crowdsourcing enrichment of multimedia collection objects with event annotations. The innovative interface combines Web technology and theory of interpretation to allow for browsing the network of data in an intuitive "infinite" fashion.

DIVE+ is result of a true inter-disciplinary collaboration between computer scientists, humanities scholars, cultural heritage professionals and interaction designers. As part of this effort, DIVE+ is also integrated in the CLARIAH[2] (Common Lab Research Infrastructure for the Arts and Humanities) research infrastructure, next to other media studies research tools, that aims at supporting the media studies researchers and scholars by providing access to digital data and tools. Research in the context of CLARIAH [2] indicates that humanities scholars need four main steps to describe their research: *exploration*, *assembly*, *analysis* and *presentation*. The DIVE+ demonstrator is able to connect to all these stages through its functionality and its integrated and interlinked datasets: *(1)* exploration: through the intrinsic nature of the tool; *(2)* assembling: finding of relevant links to additional collections; *(3)* analysis: explore, visualize and interpret the corpus through understanding the enrichment and the links between concepts in the corpus; *(4)* presentation: create informative narratives through the innovative DIVE+ interface. The overall goal of DIVE+ is to gain insights on the scalability, robustness and reusability of the digital hermeneutics approach by testing its usability in user studies with a variety of end users (both research scholars and general audiences).

2. DATA SOURCES AND ENRICHMENT

The DIVE+ 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 *four 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. The data is accessed through a SPARQL interface, on top of which an innovative and intuitive event-centric browsing interface is developed.

- *3000 Dutch news broadcasts (1920-1980)* form the Netherlands Institute for Sound and Vision (NISV)[3].
- *197,199 ANP Radio News Bulletins (1937-1955)* from the Dutch National Library (KB)[4].
- *3500 Cultural heritage objects (1950-1980)* from the Amsterdam Museum (AM)[5].
- *964 Cultural heritage objects (1950-1980)* from the Tropenmuseum (KIT) [6].

Additionally, in the DIVE+ 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), Amsterdam Museum Thesaurus, Persons list and Geo vocabulary. Thus, the collections made available are interlinked
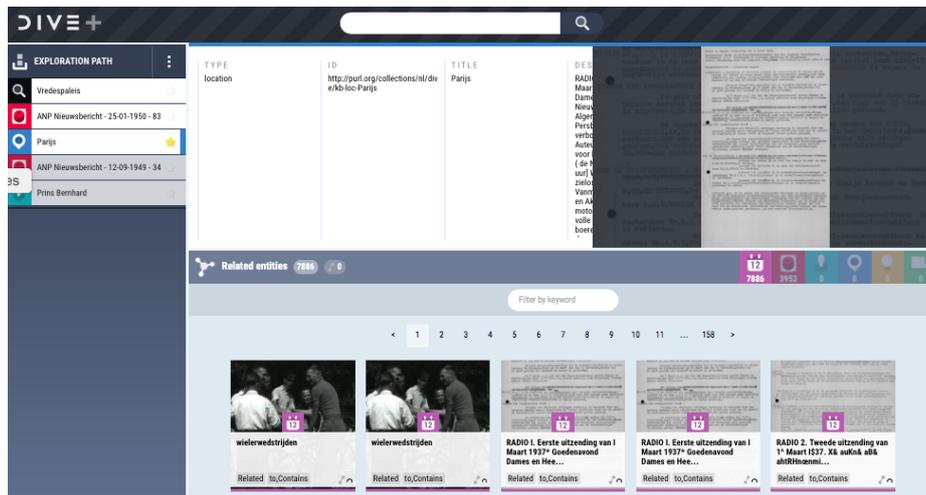
---

[1]http://diveplus.beeldengeluid.nl
[2]http://www.clariah.nl

[3]http://www.beeldengeluid.nl
[4]http://www.kb.nl,http://radiobulletins.delpher.nl/
[5]https://www.amsterdammuseum.nl
[6]http://www.opencultuurdata.nl/wiki/tropenmuseum/

Figure 1: DIVE+ Screenshot: shows a set of entities that are saved for further exploration (left side of the screen). We look into depth into the location entity "Parijs" (which is also bookmarked) - top side and then we filter based on events to get all the events that are related to this location - bottom side.

in a common linked data network of events, persons, places and concepts, which provides context for browsing and exploration of the cultural heritage objects.

The textual descriptions and descriptive metadata for all four 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[7] are consolidated to RDF. The data is modeled using the Simple Event Model (SEM) [4]. 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[8]. In the current triple store we host over 7.5 Million triples for the 210,000 Media Objects. These are annotated with 17,209 places, 95,977 actors and 199,116 event concepts.

3. USER INTERFACE

DIVE+ implements an intuitive event-centric browsing interface for browsing online heritage collections by means of an underlying linked data graph. The DIVE+ 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. During user studies with digital humanities professionals and students we tested

the usability of the DIVE+ approach. Based on the results, we further adapted the innovative DIVE+ interface to be more suitable for narrative generation and paths generation between heritage collection objects (Figure 1).

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[9] or the OpenCultuurData API[10]. 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.

For future, we plan to perform more user studies within CLARIAH, to gain insights on the scalability, robustness and reusability of the DIVE+ digital hermeneutics approach, as well as its usability for general audience.

4. 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] L. Melgar, M. Koolen, H. Huurdeman, and J. Blom. A process model of scholarly media annotation. In *CHIIR 2017, (to appear)*.

[3] 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.

[4] 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.

---

[7] http://crowdtruth.org/
[8] ClioPatria triple store data.dive.beeldengeluid.nl

[9] http://www.mediawiki.org/wiki/API:Main_page
[10] http://www.opencultuurdata.nl/api/