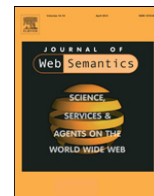




Contents lists available at ScienceDirect

# Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: [www.elsevier.com/locate/websem](http://www.elsevier.com/locate/websem)

## DIVE into the event-based browsing of linked historical media



Victor de Boer<sup>a,b,\*</sup>, Johan Oomen<sup>a</sup>, Oana Inel<sup>b</sup>, Lora Aroyo<sup>b</sup>, Elco van Staveren<sup>c</sup>,  
Werner Helmich<sup>d</sup>, Dennis de Beurs<sup>d</sup>

<sup>a</sup> Netherlands Institute for Sound and Vision, Hilversum, The Netherlands

<sup>b</sup> Department of Computer Science, VU University Amsterdam, The Netherlands

<sup>c</sup> Dutch National Library, The Hague, The Netherlands

<sup>d</sup> Frontwise, Utrecht, The Netherlands

### ARTICLE INFO

#### Article history:

Received 6 February 2015

Received in revised form

3 June 2015

Accepted 11 June 2015

Available online 18 June 2015

#### Keywords:

Digital history

Heterogeneous data cloud

Digital hermeneutics

Historical events

Crowdsourcing

### ABSTRACT

DIVE is a linked-data digital cultural heritage collection browser. It was developed to provide innovative access to heritage objects from heterogeneous collections, using historical events and narratives as the context for searching, browsing and presenting of individual and group of objects. This paper describes the DIVE Web Demonstrator.<sup>1</sup> We also discuss how the collection metadata the demonstrator uses are enriched through a hybrid workflow. The demonstrator uses semantics from these enriched collections, their vocabularies and linked data vocabularies to establish connections between the collection media objects and the events, people, locations and concepts that are depicted or associated with those objects. The innovative interface combines Web technology and theory of interpretation to allow for browsing this network of data in an intuitive “infinite” fashion. DIVE focuses to support digital humanities scholars in their online explorations and research questions.

© 2015 Elsevier B.V. All rights reserved.

## 1. Background

The Web has offered cultural heritage institutions and their public a medium, changing their traditional task from information interpreters to that of information providers [1] and collections are being made digitally available in increasing numbers. Public repositories such as Europeana and the Digital Public Library of America, for instance, offer access to tens of millions of digital artifacts from museums, archives and libraries. This urges cultural heritage institutions to rethink the access provision strategies to their collections to allow the public to interpret and contribute to their collections.

Search and browsing interfaces provide access to both professionals as well as the general public, searching for cultural heritage objects in either a single collection or in multiple collections at the same time. The traditional information access to cultural heritage

assumes that experts interpret and curate their collections in such a way that the users of their information systems perform simple or complex keyword search to come to a selection of items matching the query. However, research has shown that many users seek more *exploratory* forms of browsing [2].

Recent technical innovations, which include Linked Data, make it possible to create interactive access to cultural heritage collections not only through direct textual keyword search, but also through structured links between cultural heritage objects and related events, persons, places and concepts. In the Agora project,<sup>2</sup> browsing of cultural heritage collections through events and their links to collection object has proved successful for supporting the interpretation of end users, and thus realizing the so-called “digital hermeneutics” [3].

## 2. The DIVE demonstrator

Building on the event modeling of the Agora project, the DIVE demonstrator presented in this paper implements the event-based browsing of cultural heritage objects from two heterogeneous historical collections. Within DIVE, new interaction concepts for

\* Corresponding author at: Department of Computer Science, VU University Amsterdam, The Netherlands.

E-mail addresses: [v.de.boer@vu.nl](mailto:v.de.boer@vu.nl) (V. de Boer), [joomen@beeldengeluid.nl](mailto:joomen@beeldengeluid.nl) (J. Oomen), [oana.inel@vu.nl](mailto:oana.inel@vu.nl) (O. Inel), [lora.aroyo@vu.nl](mailto:lora.aroyo@vu.nl) (L. Aroyo), [elco.vanStaveren@kb.nl](mailto:elco.vanStaveren@kb.nl) (E. van Staveren), [werner@frontwise.com](mailto:werner@frontwise.com) (W. Helmich), [dennis@frontwise.com](mailto:dennis@frontwise.com) (D. de Beurs).

<sup>1</sup> The DIVE demonstrator is available at <http://dive.beeldengeluid.nl>.

<sup>2</sup> <http://agora.cs.vu.nl>.

events and event-based narratives have been explored and developed. We explicitly support a diversity of user groups, including Digital Humanities researchers, professional (commercial) users and the general public. The DIVE Web demonstrator is a linked-data digital cultural heritage collection browser. It was developed to provide innovative access to heritage objects from heterogeneous collections, using historical events and narratives as the context for searching, browsing and presentation of individual objects and groups of objects. DIVE provides a novel way to support digital humanities scholars in their online explorations and research questions.

The heterogeneous collections made available through the demonstrator are interlinked in a common linked data network. This interconnected network of events, persons, places and concepts provides context to the cultural heritage objects, which are represented in the same networks. Thus, the objects are contextualized with events and narratives, which is crucial for the findability and hermeneutics. Core contribution is the innovative user interface supporting information interpretation in multimedia collections through dynamic browsing experience with linked data and explicit event representation. DIVE expands on linked media browsers such as NoTube N-screen<sup>3</sup> or HyperTed<sup>4</sup> by supporting exploration of multiple types of linked media and an event-centric approach. This event-based browsing can also be found in tools such as seen.co<sup>5</sup> or Eventify.<sup>6</sup>

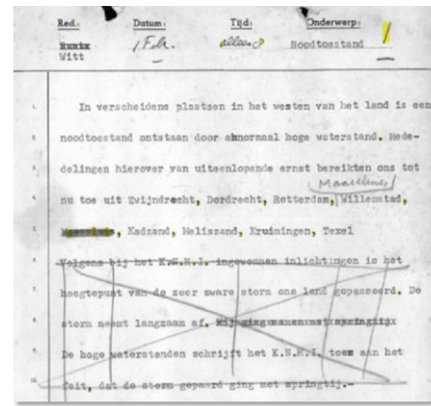
### 3. The collection data

The DIVE demonstrator allows for browsing of heterogeneous linked datasets as long as they contain media objects (such as images or videos) which are enriched through links with entities such as events, persons and places. Currently, content from two cultural heritage institutions are enriched, linked and made available. We here describe the original data and the enrichment process.

#### 3.1. Data sources

Fig. 1 shows examples of the media objects in the two collections included in the current version of the DIVE demonstrator.

- The Netherlands Institute for Sound and Vision (NISV)<sup>7</sup> archives Dutch broadcasting content, including television and radio content. A part of the NISV collection of broadcast video was published as Open Data on the Openimages.eu platform.<sup>8</sup> Within the DIVE project, a subset of this collection of 100 randomly selected videos was ingested through the OAI-PMH protocol.<sup>9</sup> This collection consists mainly of Dutch news broadcasts items from the period. These videos have a typical duration of between a 1 and 10 min. For these videos descriptive metadata is available including free-text content description.
- The Dutch National Library (KB)<sup>10</sup> provides access to a number of historical datasets. In the DIVE demonstrator, we use the KB ANP Radio News Bulletin dataset.<sup>11</sup> This dataset is made up of digitized typoscripts (radio news scripts, to be read during news broadcasts) from the period 1937–1984. These have been made public through a Web interface and API. Here, the scanned images, OCR'd content and descriptive metadata is available. The



MEDIA ITEM



Dordrecht, Holland's oldest city

License Download Share Order

Reel 3

Third and final part of a three-part film about Dordrecht in the year 1926.

less

In this part a cinematic tour along the industries of different kinds located by the deeper water on the banks. The operation of a heavy safe door is being demonstrated. Also footage of the shipping traffic on the Oude Maas and Hollands Diep with shots of the railway bridge near Moerdijk.

Fig. 1. Examples of the media objects from the two collections: the top shows one typoscript from the KB dataset. Not only the textual content but also the annotations are of interest. The second image shows a video item from the Openimages collection of NISV with the descriptive metadata. The items share a mention of a geographical location (the city of Dordrecht).

original data and metadata are available in Dutch. We ingested 2210 transcripts into the DIVE demonstrator. These were selected from roughly the same period and topics as the NISV dataset to ensure that links between the collections could be established.

#### 3.2. Data conversion and enrichment

The textual descriptions and descriptive metadata for both collections are retrieved and are converted to RDF. From the textual descriptions we extract *events* as well as places, persons and concepts linked to those events which in turn are depicted by the cultural heritage objects (videos or news bulletins). For this extraction, we employ an ensemble of methods. The DIVE system incorporates a hybrid workflow for event enrichment in video collections. In this hybrid workflow the machines and the crowd collaborate in the process of extracting relevant events and event-related concepts in video collections. In the first stage Named Entity Recognition (NER) and Event extraction tools for Dutch text are used in order to retrieve a set of relevant concepts from the video descriptions. In a second stage, crowdsourcing through the CrowdTruth platform<sup>12</sup> is employed to have human-recognized entities and to refine the results from Natural Language Processing. Section 4 presents in detail this workflow. For the extraction of the News Bulletins, we also use the results of the NER employed by

<sup>3</sup> <http://nscreen.notu.be/>.

<sup>4</sup> <http://linkedtv.eurecom.fr/Hyperted/>.

<sup>5</sup> <http://seen.co>.

<sup>6</sup> <https://eventify.it/>.

<sup>7</sup> <http://www.beeldengeluid.nl>.

<sup>8</sup> <http://openimages.eu>.

<sup>9</sup> We are currently expanding this small subset.

<sup>10</sup> <http://www.kb.nl>.

<sup>11</sup> <http://radiobulletins.delpher.nl/>.

<sup>12</sup> <http://crowdtruth.org/>.

Download English Version:

<https://daneshyari.com/en/article/558409>

Download Persian Version:

<https://daneshyari.com/article/558409>

[Daneshyari.com](https://daneshyari.com)