Linking and Visualizing Television Heritage:
The EUscreen Virtual Exhibitions and the Linked Open Data Pilot

Johan Oomen
Nederlands Instituut voor
Beeld en Geluid
Sumatralaan 45, Hilversum
The Netherlands
joomen@beeldengeluid.nl

Vassilis Tzouvaras
National Technical University of
Athens
Iroon Polytexneiou 9, 15780
Zografou
Greece
tzouvaras@image.ntua.gr

Kati Hyyppä
Noterik BV
Prins Hendrikkade 120
The Netherlands
kati@noterik.nl

ABSTRACT
The EUscreen initiative represents the European television archives and acts as a domain aggregator for Europeana, Europe's digital library, which provides access to over 20 million digitized cultural objects. The main motivation for the initiative is to provide unified access to a representative collection of television programs, secondary sources and articles, and in this way to allow students, scholars and the general public to study the history of television in its wider context. This paper explores the EUscreen activities related to novel ways to present curated content and publishing EUscreen metadata as Linked Open Data.

Categories and Subject Descriptors
H.2.4 [**Systems**]: Multimedia databases.

Keywords
Visualization, Linked Open Data, TV on the Web, Metadata Interoperability, Europeana, Linked Media

Introduction
The main motivation of the EUscreen initiative is to provide a unified online access to a representative collection of television programs, secondary sources and articles, allowing students, scholars and the general public to explore European audiovisual cultural heritage.

The multidisciplinary nature of the EUscreen project is mirrored in the composition of the socio-technical nature of the consortium; comprising of 20 collection owners, technical enablers, legal experts, educational technologists and media historians from 20 countries. EUscreen represents all major European television archives and acts as one of the key domain aggregators providing content to Europeana, the platform connecting millions and millions of cultural artifacts across Europe.

The goals of the project are to (i) develop a state-of-the-art workflow for content ingestion, (ii) define content selection and IPR management methodology (35.000 items will be made available), and (iii) design and implement a front-end that accommodates requirements from several user groups.

Close cooperation between the different stakeholders in the consortium is essential in order to reach these goals. For example, the content selection policy needs to take in to account, not only the available content, but also the needs of the user groups as well as copyright issues. The workflow also needs to accommodate to the existing metadata structures, support aggregation by Europeana[1] and provide support for multilingual access. In this paper we demonstrate the EUscreen metadata workflow focusing on the metadata ingestion, the creation of the Linked Open Data and an application that uses the data.

1. EUscreen Content and Workflow
The workflow for publishing the metadata related to the video items consists of a number of consecutive steps: the metadata ingestion, their transformation to a common reference schema, enrichment and finally publication as Linked Data. This is illustrated in Figure 1.
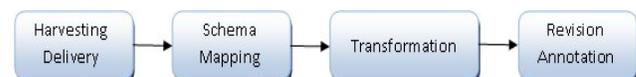

Figure 1. The ingestion workflow.

As the 20+ content providers are using different management systems and in turn different types of metadata, the need arises for interoperability. In order to achieve semantic interoperability with external web applications, a harvesting schema was implemented based on EBUCore [1], which is an established standard in the area of audiovisual metadata. An extensive evaluation of alternative standards in this area (including MPEG7, DCMI, TV Anytime) has been conducted [2] before choosing the EBUCore.

EBUCore has been purposefully designed as a minimum list of attributes that describe audio and video resources for a wide range of broadcasting applications including archives, exchange and publication. It is also a metadata schema with well-defined syntax and semantics for easier implementation. It is based on the Dublin Core to maximize interoperability with the community of Dublin Core users. EBUCore expands the list of elements originally defined in EBU Tech 3293-2001 for radio archives, also based on Dublin Core.

The MINT platform[2] (Metadata Interoperability Services) is used for the ingestion and transformation of the metadata. MINT is a web based platform for assisting the mapping of provider's existing metadata to the proposed metadata model. EUscreen

---

[1] http://www.europeana.eu

[2] http://mint.image.ece.ntua.gr

also uses MINT for enriching metadata elements, for instance to add translations of television programme titles and to add terms from the IPTC thesaurus, as a means to support multilingual retrieval.

The final step, after the transformation of the content's metadata, is the publication of the metadata as Linked Open Data [3]. Linked data aims to make data accessible, not only to humans, but also to software agents, building a semantic layer to improve and enrich their interaction.

Content Providers use the back-end tools, while the different audiences can access the content (including versatile metadata) through the EUscreen portal, which serves as the front-end, providing different ways to explore the content (Sections 2 and 3) such as advanced search and curated exhibitions. Section 4 provides more detail regarding the project's efforts related to LOD.

2. The Front-end Design

The metadata links to the media files, stored on distributed servers maintained by the project partners. Strict rules were put in place to guarantee a uniform play-out behavior of the files.

Representatives of the four primary user groups, i.e. secondary education, academic research, the general public and the cultural heritage domain were consulted in order to define the user requirements and the design of the EUscreen portal. The main challenge for the portal's front-end (Figure 2.) is to include advanced features for specific use cases without overwhelming the users with a complex interface.
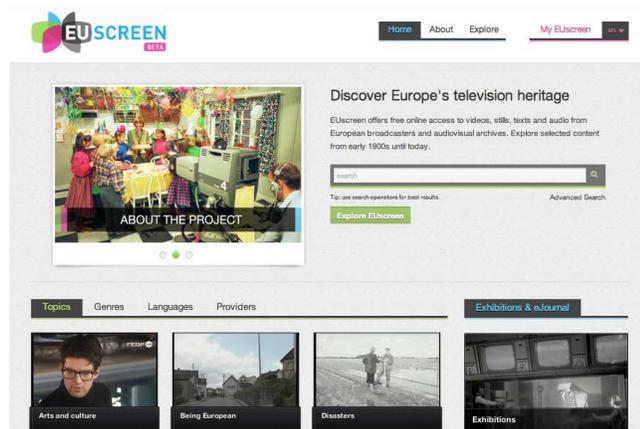

Fig. 2. EUscreen homepage.

Implementation of the front-end services is not done in the traditional way using server-side programming language like php, java or asp. Instead, it consists of 'server-less' front-end APIs where a JavaScript/Flash proxy system handles the communication with the back-end services. The resulting front-end system can be 'installed' on any plain HTML web server without additional server-side technologies. This means it can be hosted and moved to any location, or multiple locations, and that partners (Content Providers in particular) can use the APIs to integrate parts of the functionality in their own intranet and online systems using simple 'embed' mechanisms.

3. The Linked Open Data Pilot

Following the metadata format harmonisation, a Linked Open Data publication procedure has been established. This required the conversion of the harvested metadata to RDF using an expressive data model. The RDF representation of EBUcore

(http://tech.ebu.ch/lang/en/MetadataEbuCore) was used. Finally, internal and external linking to the EUscreen content has been performed and the resulting repository was made accessible through a SPARQL query endpoint. Once available, SPARQL endpoints will need to be masked by a GUI (graphical User Interface) hiding the complexity of SPARQL queries. The GUI will provide a framework for search specific to a domain of application or user profile (e.g. public access point vs. professionals or academics)

Early 2012 EUscreen launched its so-called Linked Open Data Pilot, based on the Resource Description Framework (RDF) representation of the data, and structured it using the EBU Core ontology. Aim of the pilot is to demonstrate the added value of adhering to LOD standards in the cultural heritage domain.

The data (available under the CC0 license) served by lod.euscreen.eu/resource/ is generated from the metadata EUscreen has harvested usiing the MINT ingestion platform.

We have implemented a mapping of all the harvesting schemas' elements to classes and properties of the EBU Core ontology. The EBUCore ontology is an RDF representation of the EBU Class Conceptual Data Model (CCDM). CCDM defines a structured set of audiovisual classes (e.g. groups of resources, media resources, parts, media objects but also locations, events, persons and organizations). The EBUCore ontology also defines the semantic relationships (objectProperties) between these classes as well as properties (dataProperties) characterizing these classes. A lot of the knowledge gathered in the EBU CCDM and EBUCore RDF was used to develop the W3C Media Annotation ontology[3] (W3C MAWG). Reciprocally, EBUCore RDF has implemented in a subsequent version the RDF modeling options chosen by W3C MAWG.

Therefore every provider's instance that has been transformed according to the harvesting schema, is also transformed to an RDF file. This file is then stored in a triplestore repository from where the users can consume data using SPARQL.

In addition to the original data harvested by EUscreen from its data providers, the data served at lod.euscreen.eu/resource/ includes:

- internal links to other videos that are also served by lod.euscreen.eu/resource/
- semantic enrichment done by NTUA, connecting EUscreen videos to structured representations of places and languages.

In particular, the EUscreen content has been enriched by its linking to external data sources like the DBpedia, Eurostat, Freebase and NY Times allowing for more expressive search and retrieval.

The store[4] is serving data over the HTTP protocol, using established linked data recipes, most notably the use of HTTP URIs as identifiers and entry points into data. The data can also be consumed through the SPARQL endpoint or by using the web interface of the 4store repository.

3.1 The EUscreen Timeline

As a showcase, EUscreen created a visualization that uses the LOD end-point. This visualization offers new possibilities for

---

[3] www.w3.org/2008/WebVideo/Annotations/

[4] lod.euscreen.eu/

scholars and researchers to examine the EUscreen collection on a timeline (Figure 6, and also accessible online[5]).

Above the timeline itself, a group of links is displayed, representing all of the genres and topics defined for EUscreen. Using these links, a user can filter items in the timeline, in the way that only items of the selected genres and topics are shown. When no genres or topics are selected, all items are shown.

In order to implement this demonstrator the main task was to interface with the LOD end-point hosted as defined in the Section above. To visualize the collection data in a timeline, the (JavaScript-based) SIMILE timeline[6] widget is used in combination with AJAX and a Django[7] server application of which the latter takes care of the actual communication with the SPARQL end-point to request the metadata.

4. Conclusions and future work

In this paper, we have summarized the main technical results of the EUscreen project so far, and showed initial results in the areas of visualization and Linked Open Data. In future work, we will work on:

- advanced metadata and ontology enrichment to be provided through entity extraction (e.g. names, places) from existing metadata.
- Assessing potential of time-based cross-linking with content on Europeana.

In parallel to project-based activities, we hope that the developer community takes advantage of this unique and rich resource of European television heritage. EUscreen strongly believes that providing access to their content by exploiting LOD will be an important driver for innovation.

In relation to the virtual exhibitions, we have identified two areas of work that future work will be focused on:

- Exception-based IPR. Allowing content owners to set the rights of reuse in flexible ways per collection, video or even parts of video's.
- Content-based Editing. Editing video not based on time but by adding time-based entity information and concepts so it would be possible to shift the moment of editing from production moment to consuming moment and allow for multiple on montages still reflecting the wishes/intent of the creators.

This work will be execured in the context of the three-year project EUscreenXL, starting start in 2013. We anticipate these alternative item discovery methodologies will attract new user-groups to engage with the rich body of content offered by the archives represented in the EUscreen consortium.

REFERENCES

[1] Evain, Jean Pierre (ed.) (2009) "EBU Core Metadata Set EBU", Consulted January 31, 2012. Available: http://tech.ebu.ch/docs/tech/tech3293v1_3.pdf

[2] Schreiber, G (2010). "Metadata Models, Interoperability Gaps, and Extensions to Preservation Metadata Standards"

[3] Buerger, T. & Evain, J-P. & Champin P-A. (2011), "W3C Media Annotation Working Group RDF ontology" (ma-ont) , Consulted January 31, 2012. Available: http://www.w3.org/ns/ma-ont.rdf (see also http://www.w3.org/TR/2011/PR-mediaont-10-20111129/).

---

[5] http://blog.euscreen.eu/?page_id=3117

[6] http://www.simile-widgets.org/timeline/

[7] https://www.djangoproject.com/