

Television Heritage Linked and Visualized

The EUscreen Virtual Exhibitions and the Linked Open Data Pilot

Johan Oomen

Research & Development
Netherlands Institute for Sound and Vision
Hilversum, The Netherlands
joomen@beeldengeluid.nl

Erwin Verbruggen

Research & Development
Netherlands Institute for Sound and Vision
Hilversum, The Netherlands
everbruggen@beeldengeluid.nl

Vassilis Tzouvaras

Image, Video and Multimedia Systems Lab (IVML)
National Technical University of Athens
Athens, Greece
tzouvaras@image.ntua.gr

Kati Hyypä

Noterik
Amsterdam, The Netherlands
kati@noterik.nl

Abstract—The EUscreenXL initiative represents the European television archives. It 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 EUscreenXL initiative is to aggregate a comprehensive amount of professional audiovisual content and make it accessible through Europeana. EUscreenXL builds on the EUscreen project, which provided unified access to a representative collection of television articles, and in this way allowed students, scholars and the general public to study the history of television in its wider context. This paper gives an overview of related projects that work on bringing audiovisual heritage online. It furthermore explores the EUscreen activities related to [1] novel ways of presenting curated content in virtual exhibitions and [2] publishing EUscreen metadata as Linked Open Data. Regarding the latter, it is demonstrated how available metadata can be enriched and visualized using a timeline interface.

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

I. INTRODUCTION

The main motivation of the EUscreenXL initiative is to aggregate a comprehensive amount of professional audiovisual content and make it accessible through Europeana.¹ It builds on the EUscreen project, which provides unified online access to a representative collection of television programmes, secondary sources and articles, thereby allowing students, scholars and the general public to explore Europe's audiovisual cultural heritage.

The multidisciplinary nature of the EUscreenXL project is mirrored in the socio-technical composition of the consortium. It is comprised of 29 collection owners, technical enablers, legal experts, educational technologists and media historians from 21 countries. EUscreenXL represents all major European television archives and acts as one of the key domain

aggregators providing content to Europeana, the platform connecting millions of cultural artefacts across Europe.

The goals of the project are to (i) make audiovisual content from broadcasters and audiovisual archives visible, findable and accessible across Europe, (ii) use and enhance interoperability protocols developed in its precursor EUscreen and to set up a cloud-based metadata gathering platform accessible to all relevant audiovisual content holders, (iii) maximise impact of audiovisual heritage by developing contextualisation strategies, (iv) enhance the use base of Europeana by developing and testing user engagement pilots for three end-user groups that cater for their diverse needs in order to participate and re-use material, (v) contribute to open data advocacy in the audiovisual domain in order to promote online access to cultural heritage amongst all stakeholders and (vi) build and share knowledge and raise awareness of Europeana across audiovisual archives in Europe.

As in the precursor project EUscreen, close cooperation between the different consortium stakeholders is essential to reaching these goals. The content selection policy, for example, needs to take into account the available content as well as the needs of user groups and issues related to copyright. The workflow needs to accommodate existing metadata structures, support aggregation by Europeana and provide support for multilingual access. In this paper we demonstrate the metadata workflow set up in EUscreen, focusing on metadata ingestion, the creation of Linked Open Data and an application to put these data to use.

The paper is organized as follows. The first section highlights the EUscreen project, its main objectives and metadata ingestion workflow. After discussing the related work in section two, section three presents the front-end and design of the EUscreen portal. Section four discusses the design and development of tools for creating virtual exhibitions as a novel way to navigate through the collection in a curated fashion. Section five illustrates the procedure followed for the publication of metadata as *Linked Open Data (LOD)*. Finally, a

¹ <http://www.europeana.eu>

timeline application that makes use of the LOD pilot is presented.

II. RELATED WORK

The most closely related project to EUScreenXL is the *European Film Gateway* (EFG)² and its successor, *EFG1914-1918*. Rooted in the film archiving domain, EFG provides links to cinema history related materials and contributes this information to Europeana. References to moving images on the portal point to local websites where the materials can be watched in varying formats and qualities.

Moving image archives are continuously looking for ways to bring their content online. They seek to do so to strengthen the reach of their collections without foregoing the uniqueness of their materials. The dominating power in online video, *YouTube*, is far from a perfect fit for that vision [1]. Amongst other reasons for this is its lack of support for extensive metadata models, lack of contextual information, lack of flexibility of integration with external platforms (ingest, export) and lack of advanced methods to prohibit illegal downloading. Video sharing sites such as *YouTube* do lead the way in introducing new interaction mechanisms.

In the course of 2011, different European platforms have sprung up to let users gain access to audiovisual heritage materials. The Swedish *Filmarkivet*³ launched in February 2011 with 300 short films, non-fiction films, newsreels and commercials: films that ‘reflect the transformation of Swedish society over the last century’ [2]. The main motivation for the unified access to representative programmes, secondary sources and documents with online video sources is the Dutch website *Film in Nederland*⁴ that was launched in 2011.

MAVISE holds key profile and contact data for around 40 national TV markets, 10,348 television channels and 7,901 companies. This database was developed by the Strasbourg-based *European Audiovisual Observatory* at the request of the DG Communication of the European Commission and launched in 2008. MAVISE is a tool for the television industry and parts of the database are freely accessible to the general public. MAVISE does not, however, provide access to items in respective collections.

Multiple archives are host to educational websites specifically tailored to national educational requirements. Eight different museums, institutions and audiovisual archives in the Netherlands, for example, developed *Les 2.0*, which launched in September 2010 with extensive facilities such as video editing and presentation tools.

Film and audiovisual archives have been experimenting with using the web as a means to establish communications between their collections and their users. By giving viewers access to archival content, they can subsequently add to that collection by adding both stories and content. *Wir Waren so Frei* was a project run in 2009 that amassed amateur footage from the fall of the Berlin wall. The website *Oorlog in blik*

gathered both professional and amateur film recordings to visualise World War II.

A more recent, overseas example to put the power of online crowdsourced archiving to good use, was the *Understanding 9/11* project. Launched in August, 2011, it is an online library of ‘news coverage of the events of 9/11/2001 and their aftermath as presented by U.S. and international broadcasters’ [3], with 3,000 hours of footage from a time span of seven days from 20 different international TV news channels. Additionally, select academic analysis is brought online with the collection for further contextualization. An attempt to preserve audiovisual history and actuality as it happened, was set into motion by the ‘*Archive-It*’ team, who created the Occupy Wall Street collection as an archival approach to independent media reporting.

This online back-and-forth between users and institutions fits into a wider trend in museums to move catalogues and exhibitions online and to use open source cataloguing platforms for the creation of online exhibitions. Examples of this open technology are platforms like *CollectiveAccess*, that was used for the *Wir waren so Frei* initiative, *OpenCollections* and *Omeka*, which forms the basis for the Europeana virtual exhibitions.

III. EUSCREEN CONTENT AND WORKFLOW

In collaboration with leading television historians EUScreen defined a content selection policy [4], divided into three strands:

- Historical Topics: 14 important topics in the European history in the 20th Century (70% of content);
- Comparative Virtual Exhibitions: two specially devised topics that explore particular aspects of European history (10% of content complemented with contextualizing texts and other items);
- Content Provider Virtual Exhibitions: Each content provider showcases interesting audiovisual content, supporting it with other related digital materials.

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.

As the 20+ content providers are using different management systems and in turn different types of metadata, the need for interoperability arose. In order to achieve semantic interoperability with external web applications, a harvesting schema was implemented based on EBUCore [5], 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 before choosing EBUCore [6].

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 Dublin Core in order to maximize interoperability

² <http://europeanfilmgateway.eu>

³ <http://www.filmarkivet.se>

⁴ <http://www.filminnederland.nl>

with the community of Dublin Core users. EBUCore expands the list of elements originally defined in *EBU Tech 3293-2001* for radio archives, which was also based on Dublin Core.

The MINT platform⁵ is used for the ingestion and transformation of the metadata. MINT is a web-based platform for assisting the mapping of providers' existing metadata to the proposed metadata model. EUScreen also uses MINT for enriching metadata elements, for instance to add translations of television program 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* (LOD) [7]. 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 VI provides more detail regarding the project's efforts related to LOD.

IV. 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 behaviour 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

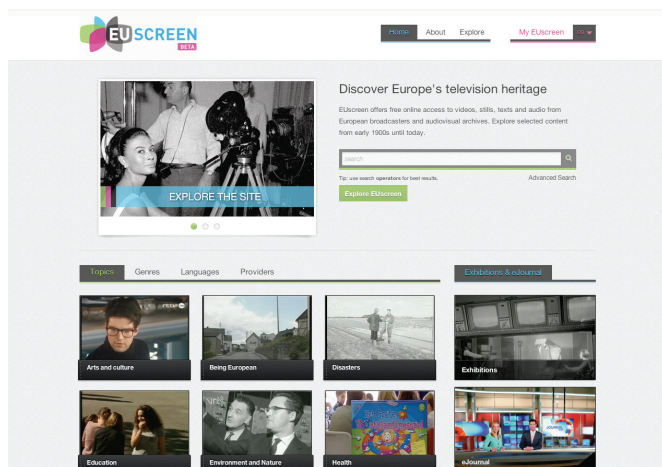


Fig. 1: EUScreen homepage

the user requirements and the design of the EUScreen portal. The main challenge for the portal's front-end (Figure 1) is to include advanced features for specific use cases without overwhelming the users with a complex interface.

Implementation of the front-end services is not done in a traditional way, using a server-side programming language like

⁵ MINT stands for Metadata Interoperability Services - <http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki>

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.

V. THE VIRTUAL EXHIBITIONS

The design and implementation of the online exhibitions started in summer 2011 and was led by Noterik⁶ and the Aalto University School of Arts, Design and Architecture⁷. The idea is to provide an online tool to allow users to create online presentations on a theme of their interest using videos and other media from EUScreen. The list of exhibitions created by academic partners and archive personnel includes cultural comparisons on topics such as fashion and food, as well as historical overviews of television's evolution or historical landmarks in specific countries.

Design workshops have constituted an important part of designing the exhibition tools. There, programmers and interface designers worked in close collaboration with scholars that curated the first exhibitions. This collaboration, amplified by further testing and feedback from EUScreen content providers, has enabled taking into account the needs and different perspectives regarding audiovisual storytelling.

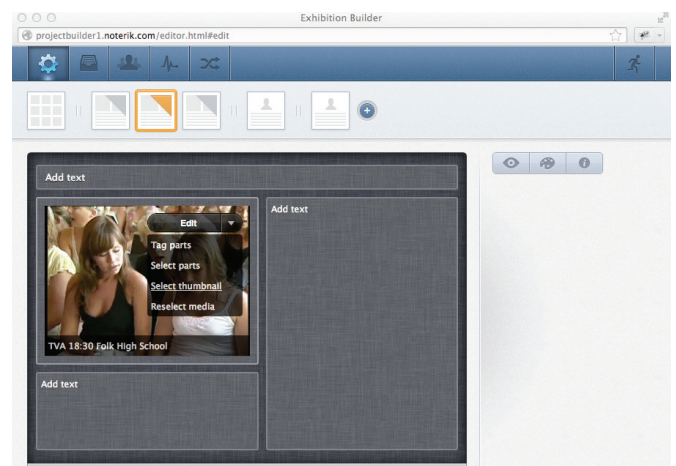


Fig. 2: The Exhibition Builder Tool

As a result, the Exhibition Builder tool has been developed, which provides a WYSIWYG style interface for creating online exhibitions (Figure 2). With this tool, it is possible to select different layouts, add media from EUScreen, and enrich it further with, for example, texts, links to external websites or direct links to parts of videos. The exhibitions are published on the EUScreen portal and can also be embedded on external websites. The exhibitions demonstrate the added value of curating audiovisual content from an online resource such as

⁶ <http://www.noterik.nl>

⁷ <http://arts.aalto.fi/en/>

EUScreen. The latest batch of exhibitions, as created by the consortium members, has been released in November 2012.

VI. THE LINKED OPEN DATA PILOT

Following the metadata format harmonization, a Linked Open Data publication procedure was established. This required the conversion of the harvested metadata to RDF using an expressive data model. Internal and external linking to the EUScreen content was performed and the resulting repository has been made accessible through a SPARQL query endpoint. Once available, SPARQL endpoints will need to be masked by a graphical user interface (GUI) that hides the complexity of SPARQL queries. The GUI will provide a framework for searches that are specific to a domain of application or to a user profile.⁸

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

The data is available under a CC0 license and generated from the metadata EUScreen has harvested using the MINT ingestion platform. In addition to the original data harvested by EUScreen from its data providers, the data served at <http://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 the NYTimes, allowing for more expressive search and retrieval. The store serves 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 store repository.

A. The EUScreen Timeline

As a showcase, EUScreen created a visualization that uses the LOD end-point. This visualization offers new possibilities for scholars and researchers to examine the EUScreen collection on a timeline and is accessible online.⁹ Above a timeline, a group of links is displayed, representing all the genres and topics defined for EUScreen. Using these links, a user can filter items in the timeline, in a 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 widget is used in

combination with AJAX and a Django server application of which the latter takes care of the actual communication with the SPARQL end-point to request the metadata.

VII. CONCLUSION

In this paper, we have summarized the main technical results of the EUScreen project 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 the 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 videos.
- 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 executed in the context of the three-year project EUScreenXL, which started in March 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] J. Farchy, "Economics of Sharing Platforms: What's Wrong with Cultural Industries?," in *The YouTube reader*, P. Snickars and P. Vonderau, Eds. Stockholm: National Library of Sweden, 2009, pp. 360–371.
- [2] Swedish Film Institute, "On filmmarkivet.se," *Filmmarkivet*, Feb-2011. [Online]. Available: <http://filmmarkivet.se/sv/English/>. [Accessed: 08-May-2012].
- [3] "Understanding 9/11: A Television News Archive," *Internet Archive*, 24-Aug-2011. [Online]. Available: <http://archive.org/details/911>. [Accessed: 08-May-2012].
- [4] L. Kaye, M. Rendina, and R. Turnock, "Content Selection and Metadata Handbook," EUScreen project deliverable D3.1.
- [5] J.-P. Evain, "EBU Core Metadata Set (EBUCore)," EBU-UER, Geneva, Specification EBU-TECH 3293, Oct. 2011.
- [6] G. Schreiber, "Metadata models, interoperability gaps and extensions to preservation metadata standards," D2, 2010.
- [7] T. Buerger, J.-P. Evain, and P.-A. Champin, "W3C Media Annotation Working Group RDF ontology," 2011.

⁸ E.g. public access point vs. professionals or academics

⁹ http://blog.euscreen.eu/?page_id=3117