PUBLISHING EUROPE'S TELEVISION HERITAGE ON THE WEB: THE EUSCREEN PROJECT

Johan Oomen, Netherlands Institute for Sound and Vision, Hilversum, Netherlands and Vassilis Tzouvaras, National Technical University of Athens, Greece

1.   Introduction

The EUscreen project represents the European television archives and acts as a domain aggregator for Europeana, Europe's digital library. The main motivation for it 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. The main goals of EUscreen are to (i) develop a state-of-the-art workflow for content ingestion, (ii) define content selection and IPR management methodology, and (iii) provide a front-end that accommodates requirements from several user groups.

Providing access to large integrated digital collections of cultural heritage objects is a challenging task. Multiple initiatives exist in different domains. For example, Europeana[1] manages a state-of-the art technical infrastructure to manage the ingestion and management of data from a wide variety of content providers. It aims to give access to Europe's entire digitised cultural heritage by 2025. Europeana focuses on two main tasks (i) to act as a central index, aggregating and harmonising metadata following a common data model (Isaac, 2010) and (ii) to provide persistent links to content hosted by trusted sources. The portal currently provides access to 15 million objects, primarily books and photographs; audiovisual collections are underrepresented. However, recent analysis of query logs from the Europeana portal indicated users have a special interest in this type of content.

Television content is regarded a vital component of Europe's heritage, collective memory and identity — all our yesterdays — but it remains difficult to access. Even more than with the museum and library collections, the dealing with copyrights, encoding standards, costs for digitization and storage makes the process of its aggregated and contextualized publishing on the Web extra challenging.

2.   Context and motivation

The main motivation for our work is to overcome the current barriers and provide *unified access* to a representative collection of television programs, secondary sources and articles, to allow access to students, scholars and the general public. The multidisciplinary nature of the EUscreen project is mirrored in the composition of the socio-technical nature of the consortium; comprising 20 collection owners plus technical enablers, legal experts, educational technologists and media historians of 20 countries. EUscreen represents all major European television archives and acts as one of the key domain aggregators providing content to Europeana.

Several public reports on our work can be downloaded from the project blog[2]. This paper reports on the results of the work performed over the past one and a half years, leading up to the launch of the first version of the portal. Notably, we zoom in on the interplay between user requirements, technical possibilities and societal issues, including intellectual property rights. We will show how EUscreen contributes to a so-called 'Cultural Commonwealth' (American Council of Learned Societies Commission on Cyberinfrastructure, 2006) that emerges by bringing content from memory institutions and the knowledge of its heterogeneous constituency together.

Conceptually, EUscreen is built on the notion that knowledge is created through conversation

---

1   http://www.europeana.eu
2   http://www.euscreen.eu

(Scott, 2001). Hence, ample attention is given to investigating how to stimulate and capture knowledge of its users. Combining organizational, expert and amateur contributions is a very timely topic in the heritage domain, requiring investigation of the technical, organizational and legal specificities. The success of EUscreen inherently depends on the adaptation of an approach that takes the full scope of these factors into consideration.

3.   EUscreen: television on the web

The main goals of EUscreen 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) provide a front-end that accommodates requirements from several user groups.

To reach these goals, close cooperation between the different stakeholders in the consortium is essential. For example, the selection policy needs to take in to account the available content, wishes from media historians and the copyright situation. The workflow will need to study the existing metadata structures, should support aggregation by Europeana and provide support for multilingual access.

3.1. Content ingestion workflow

The technical standards enabling *interoperability* form an important dimension of the technical achievements. In order to achieve semantic interoperability, a common automatic interpretation of the meaning of the exchanged information is needed, i.e. the ability to automatically process the information in a machine-understandable manner. The first step of achieving a certain level of common understanding is a representation language that exchanges the formal semantics of the information. Then, systems that understand these semantics can process the information and provide web services like searching and retrieval.

Many different metadata schemas, or, in a broader sense, sets of elements of information about resources, are being used in this domain, across a variety of technical environments and scientific disciplines. EUscreen has developed an ingestion mechanism providing a user friendly environment that allows for the extraction and presentation of all relevant and statistical information concerning input metadata together with an intuitive mapping service that uses the EUscreen Metadata schema, and provides all the functionality and documentation required for the providers to define their crosswalks. The workflow (See Figure 1) consists of four phases, each responsible for specific services to ensure the quality of the ingestion process:
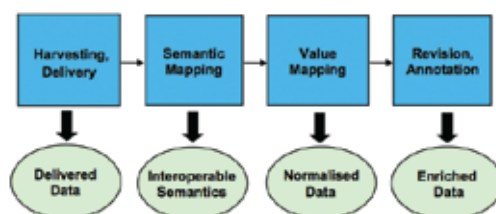

Figure 1. Content ingestion workflow

In order to achieve semantic interoperability with external web applications, EUscreen metadata are exported in EBUcore (Evain, 2009), which is an established standard in the area of audiovisual metadata. The metadata is stored in RDF format to improve the search functionality and enable the alignment with external resources. Finally, EUscreen has created a SKOS multilingual thesaurus (15 languages) based on the subject terms of the IPTC standard [International Press Telecommunications Council http://www.iptc.org/site/Home/] and the geographical places of GeoNames (http://www.geonames.org/). The thesaurus supports multilingual retrieval services and links to open data resources that could

be used for enrichment and to contextualise the collection.

3.2. Content selection

A totally different challenge is the selection of a core collection of 35,000 objects from these vast collections. In collaboration with leading television historians EUscreen has defined a content selection policy, divided into three strands:

1. *Historical Topics:* 14 important topics in the history of Europe in the 20th Century (70% of content);
2. *Comparative Virtual Exhibitions:* two specially devised topics that explore more specialised aspects of European history in a more comparative manner (10% of content — include documents, stills, articles);
3. *Content Provider Virtual Exhibitions:* each content provider selects content supported with other digital materials and textual information on subjects or topics of their own choosing (20 % of content).
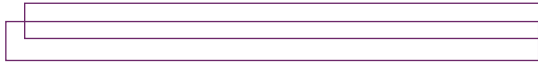
3.3. Front-end

Representatives of the four primary user groups, e.g. secondary education, academic research, the general public and the cultural heritage domain were consulted in order to define user requirements (Tzouvaras, 2010). The main challenge for the portal's front-end is to include advanced features for specific use cases without overwhelming the users with a complex interface. The Helsinki University of Arts and Design adapted a component-based conceptual model that accommodates this requirement (See Figure 2).

Figure 2. Homepage design

The Web platform also includes interactive web-based features, such as the creation of so-

called virtual exhibits and playlists that assists the users to appropriate this rich resource in a meaningful way.

4.   Future work

The first version of the portal was launched in March 2011, which is being followed by a period of extensive evaluations with end-users. Also, the selection policy will be reviewed. Outcomes of this process will form the basis of the design of the second release, scheduled for early 2012. We anticipate that the participatory features of the portal will be extended and we can further enhance the portal creating additional links to datasets in Linked Data format.

Acknowledgements

References

Isaac, Antoine (ed.). 2010. *Europeana Data Model Primer*. Europeana v1.0 Technical report.

American Council of Learned Societies Commission on Cyberinfrastructure. 2006. *Our Cultural Commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences.* American Council of Learned Societies.

Scott, B. 2001. 'Gordon Park's conversation theory: a domain independent constructivist model of human knowing'. *In: Foundations of Science* 6(4):343-360.

Evain, Jean Pierre (ed.) 2009. *EBU Core Metadata Set.* EBU

Tzouvaras, Vassilis. 2010. D4.1 - Functional Specifications & Portal Architecture.