Linking Europe's Television Heritage

Nikolaos Simou, NTUA, Greece

www.image.ece.ntua.gr/~nsimou/

Jean Pierre Evain, Metadata and Workflow Processes Groups EBU, Switzerland

http://www.linkedin.com/profile/view?id=2005927&trk=tab_pro

Vassilis Tzouvaras, NTUA, Greece

www.image.ece.ntua.gr/~tzouvaras/

Marco Rendina, Rome Research Consortium, Italy

http://it.linkedin.com/in/marcorendina

Nasos Drosopoulos, NTUA, Greece

www.image.ece.ntua.gr/~ndroso/

Johan Oomen, Netherlands Institute for Sound and Vision, Netherlands

http://nl.linkedin.com/in/johanoomen

Abstract

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. In this paper, we present the various issues that were addressed, in order to accomplish such a challenging task.

Keywords: Metadata Ingestion, Aggregation,TV on the Web, Linked Data.

1. Introduction

Digital evolution of the Cultural Heritage field has accelerated rapidly in the past few years. Massive digitization and annotation activities are in progress all over Europe and the world, following the early developments at the European level and the “Lund principles” (European Council, 2001; European Commission, 2002a:44). Furthermore, the strong involvement of companies like Google, together with the positive reaction and increasing support of the European Union, have led to a variety of, rather converging, actions towards multimodal and multimedia cultural content generation from all possible sources (i.e. galleries, libraries, archives, museums, audiovisual archives etc.).

The creation and evolution of Europeana (http://www.europeana.eu), as a unique point of access to European Cultural Heritage, has been one of the major achievements of these efforts. At the moment, more than 20 million objects, expressing the European cultural richness, are accessible through the Europeana portal, and it is expected that this number will be doubled within the next five years. The Europeana portal currently provides access to various cultural objects and their digital representations, of which the majority is text or images; audiovisual collections are underrepresented. However, recent analysis of query logs from the Europeana portal indicated that users have a special interest for 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 make the process of its aggregated and contextualized publishing on the Web extra challenging.

EUscreen project aims at the creation of a representative collection of television programs, secondary sources and articles permitting in this way access to students, scholars and the general public. However, providing access to large integrated digital collections of cultural heritage objects is a challenging task involving the resolution of various issues. Firstly, the aggregation of metadata together with a harmonizing process – since different content providers adopt different types of models – must be considered. After that, the metadata must be made available to the public in a consistent way, not only offering a user friendly navigation and preview but also allowing their consumption and re-use in a machine understandable manner.

In this paper, we present the workflows and respective tools used for the ingestion and manipulation of Europe’s Television Heritage content as well as the methodology adopted for its publication as Linked Data. Specifically the overall workflow consists of three main steps, the metadata ingestion, their transformation to a common reference schema, and finally their publication as Linked Data.

Firstly, the various content providers implied different management systems and in turn different types of metadata, arising the need for interoperability. In order to achieve semantic interoperability with external web applications, a harvesting schema was implemented based on EBUCore (Evain, 2009), which is an established standard in the area of audiovisual metadata. An extensive evaluation of alternative standards in this area (MPEG7, DCMI, TV Anytime) has been conducted (Schreiber, 2010) 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. MINT (Metadata Interoperability Services - *http://mint.image.ece.ntua.gr/*) was 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.

The next step, after the transformation of the content's metadata, was their publication as Linked Data. Linked Data aim to make data accessible not only to humans but also to software agents, building in that way a semantic layer to improve and enrich of their interaction. In order to achieve this objective, the conversion of the harvested metadata to RDF using an expressive data model was required, and in our case EBUCore ontology was the most appropriate to guide this semantic transformation. Finally, internal and external linking of the EUscreen content has been performed and the resulting repository can be accessed through its SPARQL endpoint.

The rest of the paper is organized as follows. The first section highlights the EUscreen project and its main objectives. The next two sections present the MINT platform that was used for the aggregation and transformation of the metadata, and the procedure followed for their publication as Linked Data respectively. Finally, we conclude with the discussion concerning additional and more expressive retrieval capabilities offered by taking advantage of the content's publication as Linked Data.

2. The EUscreen Project

The EUscreen project aims to promote the use of television content to explore Europe's rich and diverse cultural history. Its main objective is to create access to over 30,000 items of programme content and information, and by developing a number of interactive functionalities and dynamic links with Europeana, to prove valuable to the widest range of cultural, educational and recreational users. 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 of 20 countries.

EUscreen contributes to a so-called "Cultural Commonwealth" (Welshons, 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 (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.

In collaboration with leading television historians EUscreen has defined a content selection policy (Kaye, 2011), divided into three strands:

1. Historical Topics: 14 important topics in history of Europe in the 20th Century (70% of content);
2. Comparative Virtual Exhibitions: two specially devised topics that explore more specialized 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).

EUscreen has written a set of guidelines regarding management of intellectual property rights. The copyright situation of each and every item is investigated prior to uploading.

3. Metadata Aggregation and Transformation

In this section we present the MINT services that were used for the metadata aggregation and transformation. We mainly focus on its aggregation workflow - that is adopted by the

EUscreen project - and its mapping editor that facilitates the mapping of providers' data to the EBU based target schema.

MINT Platform

MINT is an open source, web based platform for the ingestion, mapping and transformation of metadata records. Interoperability is achieved through the use of well defined metadata models – in the way tha EBUCore was used for the EUscreen case - and the alignment of the providers' records to their requirements. The metadata ingestion workflow, as illustrated in Figure 1, consists of four main procedures.

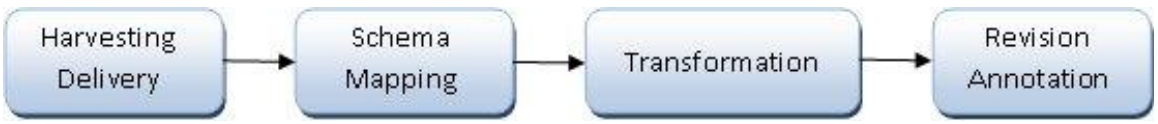

Figure 1: Ingestion Workflow

More specifically, the platform offers a user and organization management system that allows the deployment and operation of different aggregation schemes with corresponding user roles and access rights. Registered users can start by uploading their metadata records in XML or CSV serialization, using the HTTP, FTP and OAI-PMH protocols. Users can also directly upload and validate records in a range of supported metadata standards (XSD). XML records are stored and indexed for statistics, previews, access from the mapping tool and subsequent services. Handling of metadata records includes indexing, retrieval, update and transformation of XML files and records. XML processors are used for validation and transformation tasks as well as for the visualization of XML and XSLT.

The next step is the mapping procedure for which MINT uses a visual mapping editor for the XSL language. Mapping is performed through drag-and-drop and input operations which are translated to the corresponding code. The editor visualizes the input and target XSDs, providing access and navigation of the structure and data of the input schema, and the structure, documentation and restrictions of the target one. Mappings can be applied to ingested records, edited, downloaded and shared as templates.

During the third step, users can transform their selected collections using complete and validated mappings in order to publish them in available target schemas for the required aggregation and remediation steps. Preview interfaces present the steps of the aggregation such as the current input xml record, the XSLT code of mappings, the transformed record in the target schema, subsequent transformations from the target schema to other models of interest (e.g. Europeana's metadata schema), and available html renderings of each xml record.

Finally, the last step is the Revision/Annotation procedure that enables the addition and correction of annotations, the editing of single or group of items in order to assign metadata not available in the original context and, further transformations and quality control checks according to the aggregation guidelines and scope (e.g. for URLs).

Mapping Editor

Metadata mapping is the crucial step of the ingestion procedure. It formalizes the notion of a metadata crosswalk, hiding the technical details and permitting semantic equivalences to emerge as the centrepiece. It involves a user-friendly graphical environment (Figure 2 shows an example mapping opened in the editor) where interoperability is achieved by guiding users in the creation of mappings between input and target elements. User imports are not required to include the respective schema declaration, while the records can be uploaded as XML or CSV files. User's mapping actions are expressed through XSLT stylesheets, i.e. a well-formed XML document conforming to the namespaces in XML recommendation. XSLT stylesheets are stored and can be applied to any user data, exported and published as a well-defined, machine understandable crosswalk and, shared with other users to act as template for their mapping needs.

Figure 2: Screenshot of the mapping editor

The structure that corresponds to a user's specific import is visualized in the mapping interface as an interactive tree that appears on the left hand side of the editor. The tree represents the snapshot of the XML schema that is used as input for the mapping process. The user is able to navigate and access element statistics for the specific import while the set of elements that have to be mapped can be limited to those that are actually populated. The aim is to accelerate the actual work, especially for the non-expert user, and to help overcome expected inconsistencies between schema declaration and actual usage.

On the right hand side, buttons correspond to high-level elements of the target schema and are used to access their corresponding sub-elements. These are visualized on the middle part of the screen as a tree structure of embedded boxes, representing the internal structure of the complex element. The user is able to interact with this structure by clicking to collapse and expand every embedded box that represents an element, along with all relevant information (attributes, annotations) defined in the XML schema document. To perform an actual (one to one) mapping between the input and the target schema, a user has to simply drag a source element from the left and drop it on the respective target in the middle.

The user interface of the mapping editor is schema aware regarding the target data model and enables or restricts certain operations accordingly, based on constraints for elements in the target XSD. For example, when an element can be repeated then an appropriate button appears to indicate and implement its duplication. Several advanced mapping features of the language are accessible to the user through actions on the interface, including:

- String manipulation functions for input elements.
- m-1 mappings with the option between concatenation and element repetition.
- Structural element mappings.
- Constant or controlled value assignment.
- Conditional mappings (with a complex condition editor).
- Value mappings editor (for input and target element value lists).

4. EUscreen Linked Open Data pilot

In this section we present the steps followed for the publication of the EUscreen content as Linked Data. In the case of EUscreen we have adopted the Linked Open Data principles for the deployment of the pilot (reference). We first make a short introduction to Linked Data and their basic principles. This is followed by explaining the production of the RDF instances from the aggregated and transformed to the EBUCore based schema metadata (XML to RDF), together with the way that were linked to external sources.

Linked Data Principles

During the last few years the Web has evolved from a global information space of linked documents to one where both documents and data are linked. This evolution has resulted in a set of best practices for publishing and connecting structured data on the Web known as Linked Data. In few words, Linked Data is simply about establishing typed relations between web data from a variety of sources like. These may be as diverse as databases maintained by two organizations in different geographical locations, or simply heterogeneous systems within one organization that, historically, have not easily interoperated at the data level. Technically, Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets (Bizer & Heath & Berners-Lee, 2009).

The main difference of the hypertext Web and Linked Data is that the first is based on HTML (HyperText Markup Language) documents connected by untyped hyperlinks while on the other hand Linked Data relies on documents containing data in RDF (Resource Description Framework) format (Klyne & Carroll, 2004). However, rather than simply connecting these documents, Linked Data uses RDF to make typed statements that link arbitrary things in the world. The result, or as widely known the Web of Data, may more accurately be described as a web of things in the world, described by data on the Web. Berners-Lee (2006) outlined a set of 'rules' for publishing data on the Web in a way that all published data becomes part of a single global data space:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things

These have become known as the 'Linked Data principles', and provide a basic recipe for publishing and connecting data using the infrastructure of the Web while adhering to its architecture and standards.

Semantic Representation of the EUscreen content

For instantiating the EUscreen data as Linked Data resources, a machine readable representation in RDF was necessary (a procedure also known as RDFization in the semantic web community) beyond direct XML to RDF conversion. This process is not trivial like many people mistakenly may consider, misunderstanding the XML serialization of RDF. To fully understand the reason why this process is important we must pinpoint the difference of representing cultural content in XML and RDF. In our case XML is used for collecting the metadata about video content, while RDF is employed to make statements about resources (in particular Web resources) in the form of subject-predicate-object expressions (so called triples). Therefore, in contrast to the XML transformation during which an XML document is transformed to another XML document of different structure, during the RDFization process the things described in the XML document have to be firstly identified, together with the statements about them, before proceeding to the instantiation of the RDF document.

RDF provides a generic, abstract data model for describing resources using subject, predicate, object triples. However, it does not provide any domain-specific terms for describing classes of things in the world and how they relate to each other. This function is served by taxonomies, vocabularies and ontologies expressed in SKOS (Simple Knowledge Organization System) (Miles & Bechhofer, 2009), RDFS (the RDF Vocabulary Description Language, also known as RDF Schema) (Brickley & Guha, 2004) and OWL (the Web Ontology Language) (McGuinness & Harmelen, 2004). Hence, a decision that was made in accordance with the things described in the EBUCore homogenized XML documents was the selection of the EBUCore ontology (Buerger & Evain & Champin, 2011) as the vocabulary used for the RDF representation.

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 (W3C MAWG). Reciprocally, EBUCore RDF has implemented in a subsequent version the RDF modeling options chosen by W3C MAWG.

The next step, after the selection of the appropriate vocabularies for the RDF representation of the EUscreen content, was the creation of resources for the described things. In other words, to fulfill the first principle of Linked Data that states the use of URIS for things. There are various guidelines for creating cool URIs for the semantic web (Sauermann & Cyganiak, 2008; Berners-Lee, 1998) and the two basic characteristics they must have are to be unique for every item, and consistent. According to these guidelines every entity represented in our data set leads to the minting of at least three URIs

- a URI for the real-world object itself
- a URI for a related information resource that describes the real-world object and has an HTML representation (dereferencable)
- a URI for a related information resource that describes the real-world object and has an RDF/XML representation

For ensuring the uniqueness of the URIs, web resources are served under a domain administered by the project (*lod.euscreeen.eu*) and the assigned unique identifier of the item is part of the URI. The corresponding set of URIs for an example EUscreen item is shown below.

- *http://lod.euscreen.eu/resource/EUS_55F569268ACA42B186682960875F862B*
- *http://www.euscreen.eu/play.html?id=EUS_55F569268ACA42B186682960875F862B*
- *http://lod.euscreen.eu/data/EUS_55F569268ACA42B186682960875F862B*

After specifying the way that the URIs are created, the things described were identified together with the appropriate EBU Core classes and properties that would be used for their representation in RDF. More specifically, the type of video is specified in the XML document and it can be either a part of a programme or the whole programme. Depending on this information the resource created for the video can either be an instance of the EBUCore class Part (i.e. one of several media fragments -audio, video, data- that composes an audiovisual media resource; in other ontologies fragment is often referred to e.g. as a 'part' or 'segment') or a MediaResource itself. The additional characteristics of the video resources were represented in RDF by using EBUCore properties having as range either typed literals (e.g. original title was represented by ebucore:originalTile) or in some cases other internal resources (e.g. for each video provider a new resource was made that is an instance of ebucore:Agent). The complete set of properties and classes used for the mapping of all the harvesting schema's elements can be found at *https://docs.google.com/spreadsheet/ccc?key=0Akruw5a0_oaLdEQyMl85NVQxZ2lmT00wcVU4ZVRJZ0E&hl=en_US#gid=3*

Finally, another recommendation that is very important and has to be considered during Linked Data publication is ownership of resources and provenance of information. Therefore, for every RDF representation of an item provenance metadata has been published that include the publication date and the creator and consumers can track the origin of particular data fragments.

Linking of EUscreen resources

As mentioned before, Linked Data is simply about using the Web to create typed links between data from different sources, therefore after the RDF representation of the EUscreen content, links to other resources had to be created. There are two kinds of links, internal and external RDF links. Internal RDF links connect resources within a single Linked Data source. Thus, the subject and object URIs are in the same namespace. External RDF links connect resources that are served by different Linked Data sources. The subject and object URIs of external RDF links are in different namespaces. External RDF links are crucial for the Web of Data as they are the glue that connects data islands into a global, interconnected data space (Heath & Bizer, 2011).

For the case of internal linking, specific elements of the harvesting schema that relate items were used. For example, the value of the harvesting schema's element isRelatedToItem is an EUscreen item identifier. Respectively, in the RDF representation the EBU Core property isRelatedTo was used having as range the resource of the specific item. Furthermore, additional internal linking was implemented for the countries, the actors and the organizations. The specificity of this information is the fact that it is shared among the dataset. In other words a country can be the location of production of more than one video item. Therefore new resources have been created for these elements values without any identifier – and only by using their name – since those are already unique. In that case of the Netherlands the shared resource is *http://lod.euscreen.eu/resource/Netherlands*. Therefore can be used the object of a triple that has as predicate the EBU Core property createdIn and subject the video resource.

The resources implemented for the countries were also externally linked, since information about countries is served by many data sources. For the creation of external links DBpedia (*http://dbpedia.org/About*) has been used. The names of the local dataset countries were compared using SPARQL (Prud'hommeaux & Seaborne, 2008) to names of the countries resources served by DBpedia. After the establishment of a link to DBpedia additional linked data resources can be discovered by using SPARQL. In that way we have ended up with external links to all the datasources (Freebase, Dbpedia, Eurostat, NYTimes) that DBpedia is linked. In addition to these links, new external links were extracted from the video summaries by using DBpedia spotlight, a tool that can extract resources from free text (*http://dbpedia.org/spotlight*). In the summary description of a video quite often names of persons are mentioned that either participates in the video or the video involves them in a way. By the use of spotlight resources for such cases were extracted, since they can prove to provide very useful additional information about the video and therefore its semantic retrieval.

Deployment of Linked Data Pilot

So far we have described the main issues regarding the transformation of the harvested and homogenized XML items to RDF and their internal and external linking. However, in order to fulfill the 4 main Linked Data principles some additional actions had to be considered. Firstly, both the machine (RDF) and the human (HTML) understandable information (a detailed description of the HTML representation of the items, that is given through the EUscreen portal, is given in the next subsection) had to be served. For our case, having produced static RDF files, the best practice was to directly upload the files to the server. Another important issue when publishing Linked Data is to offer a content negotiation mechanism (Fielding, 1999), the basic idea of which is that HTTP clients send HTTP headers with each request to indicate what kinds of documents they prefer. The following figure illustrates the result of the VAPOR Linked Data validator (*http://vapour.sourceforge.net/*) by presenting the dereferencing procedure for the EUscreen data.

Figure 3: The VAPOR result for EUscreen data.

Finally, we have uploaded the data to 4store (*http://4store.org/*) - a purpose built database- in order to provide SPARQL access to the data making their consumption easier. In that way the data can also be consumed through the SPARQL endpoint (*http://oreo.image.ece.ntua.gr:10999/sparql/*) or by using the web interface of the 4store repository (*http://oreo.image.ece.ntua.gr:10999/test/*).

EUscreen portal

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 and design front-end functionality. The main challenge for the portal's front-end is to include advanced features for specific use cases without overwhelming the users with complex interfaces. The Helsinki University of Arts and Design adapted a component-based conceptual model that accommodates this requirement.

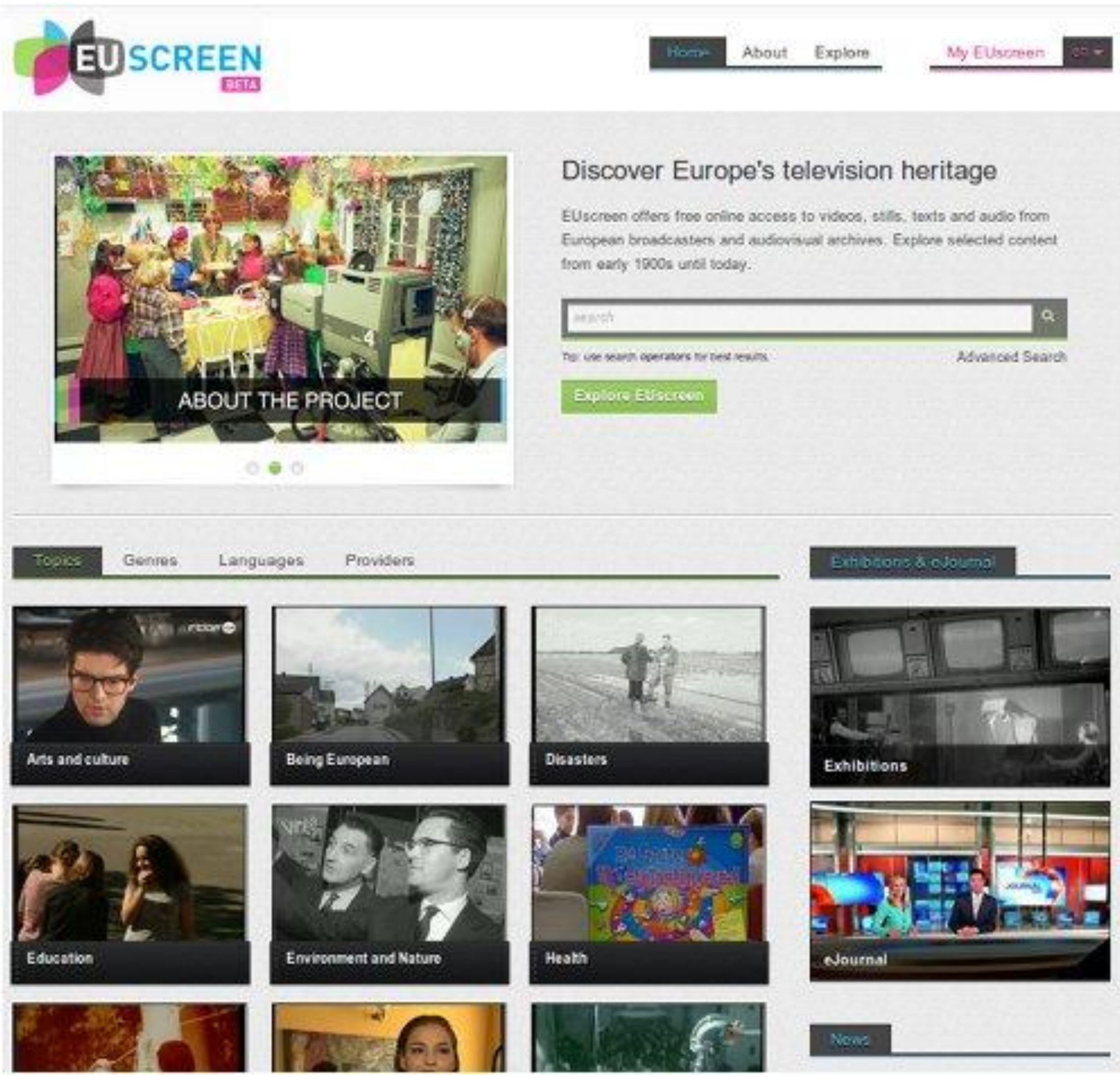

Figure 4: The EUscreen Portal

Implementation of the front-end services is not done in the traditional way using serverside programming language like php, java or asp. EUscreen implemented a 'server-less' front-end APIs where a javascript/flash proxy system handles the communication with the back-end services. The result is a front-end system that can be 'installed' on any plain html web server without any need for server-side technologies. This means it can be hosted and moved to any location or multiple locations. It also means partners can use these APIs to integrate parts of the functionality in their own intranet and internet systems using simple 'embed' ideas. This method is gaining more ground, for example companies like Google who provides these types of APIs for services like Google Maps.

5. Conclusion

In this paper, we presented the workflows and respective tools used for the ingestion and manipulation of Europe's Television Heritage content as well as the methodology adopted for its publication as Linked Data. The main benefit by the publication of the EUscreen content as Linked Data is that the data can be easily consumed, making in that way, the implementation of various applications much simpler. Moreover, 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. For example, in the case of using only EUscreen information, one can perform a query for "videos that were created in the Netherlands together with their creators". Such type of queries are supported at the moment by the portal by filtering the provider.

However, since the Netherlands is a resource that is externally linked to Eurostat which serves various statistics, one can also exploit this relevant data within the multitude of the available content by performing a query of the form "health related videos, along with the death rate for the country of broadcast". Similarly, by using the DBpedia resources about people and the classification hey offer, allow to perform queries such as "videos that are about painters".

6. References

Berners-Lee, T. (1998) "Cool uris don't change", Consulted January 31, 2012. Available: *http://www.w3.org/Provider/Style/URI*

Berners-Lee, T. (2006). "Linked Data - Design Issues", Consulted January 31, 2012. Available: *http://www.w3.org/DesignIssues/LinkedData.htm*

Brickley, D. & Guha, R. V.(2004) "RDF Vocabulary Description Language 1.0: RDF Schema - W3C Recommendation", Consulted January 31, 2012. Available: *http://www.w3.org/TR/rdf-schema/*

Bizer, C. & Heath, T. & Berners-Lee, T. (2009) “Linked Data - The Story So Far”, International Journal on Semantic Web and Information Systems, Volume: 5, Issue: 3, Publisher: Elsevier, Pages: 1-22 ISSN: 15526283, DOI: 10.4018/jswis.2009081901

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/*)

European Council (2001) “European Content in Global Networks: Coordination Mechanisms for Digitisation Programmes (Lund Principles)”, Consulted January 31, 2012. Available:*ftp://ftp.cordis.lu/pub/ist/docs/digicult/lund_principles-en.pdf*

European Commission (2002a) “Technological Landscapes for Tomorrow’s Cultural Economy: Unlocking the Value of Cultural Heritage”, DigiCULT Report, Luxembourg: Official Publications of European Communities, Consulted January 31, 2012. Available: *http://www.digicult.info/pages/report.php*

Evain, Jean Pierre (ed.) (2009) “EBU Core Metadata Set EBU”, Consulted January 31, 2012. Available: *http://tech.ebu.ch/docs/tech/tech3293v1_3.pdf*

Fielding, R. (1999) “Hypertext transfer protocol http/1.1”, Consulted January 31, 2012. Available: *http://www.w3.org/Protocols/rfc2616/rfc2616.html*

Heath, T. & Bizer, C (2011) “Linked Data: Evolving the Web into a Global Data Space” (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool DOI: 10.2200/S00334ED1V01Y201102WBE001 ISBN: 9781608454303 (paperback) ISBN: 9781608454310 (ebook).

Kaye, L. (2011) “Content Selection and Metadata Handbook” , Consulted January 31, 2012. Available: *http://blog.euscreen.eu/wp-content/uploads/2010/10/Content-Selection-and-Metadata-Handbook_public.pdf*

Klyne, G. & Carroll, J. (2004). “Resource Description Framework (RDF): Concepts and Abstract Syntax” - W3C Recommendation. Retrieved June 14, 2009, Consulted January 31, 2012. Available: *http://www.w3.org/TR/rdfconcepts/*

McGuinness, D. L. & Harmelen, F.V (2004) “OWL Web Ontology Language Overview - W3C Recommendation”, Consulted January 31, 2012. Available: *http://www.w3.org/TR/2004/REC-owl-features-20040210/*

Miles, A. & Bechhofer, S (2009). “Skos simple knowledge organization system”, Consulted January 31, 2012. Available: *http://www.w3.org/TR/skos-reference/*

Prud’hommeaux, E. & Seaborne, A. (2008) “SPARQL Query Language for RDF - W3C Recommendation”, Consulted January 31, 2012. Available: *http://www.w3.org/TR/rdf-sparql-query/*

Sauermann, L. & Cyganiak, R (2008). “Cool uris for the semantic web - w3c interest group note”, Consulted January 31, 2012. Available: *http://www.w3.org/TR/cooluris/*

Schreiber, G (2010). “Metadata Models, Interoperability Gaps, and Extensions to Preservation Metadata Standards”

Scott, B. (2001). “Gordon Park's conversation theory: a domain independent constructivist model of human knowing”. In: Foundations of Science 6(4):343-360. National Centre for Biotechnology Information, Consulted January 31, 2012. Available: *http://www.ncbi.nlm.nih.gov*

Welshons, M. (2006). "Our Cultural Commonwealth" The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. Connections, December 15, 2006. , Consulted January 31, 2012. Available: *http://cnx.org/content/col10391/1.2/*.