

# Interlinking Media Archives with the Web of Data

## Semantic inline annotation of online content

Dietmar Glachs, Sebastian Schaffert, Christoph Bauer

SalzburgResearch Forschungsgesellschaft m.b.H., Salzburg, Austria  
{dietmar.glachs,sebastian.schaffert}@salzburgresearch.at  
Österreichischer Rundfunk, Wien, Österreich  
christoph.bauer@salzburgresearch.at

**Abstract.** Today's enterprises heavily rely upon accurate, consistent, and timely access to data. However, company data is typically scattered across multiple databases and file shares in a multitude of forms and versions. Moreover, an increasing amount of valuable background information is available outside the companies' influence and control. This situation is typical for many enterprise information integration scenarios, also in Austria's largest broadcasting media archive. Our demonstration argues for an information integration approach that uses semantic web principles to interlink archival media content of the Austrian Broadcasting Corporation (ORF) with the web of data and with internal knowledge resources to facilitate semantic search and to increase the user experience of browsing and discovering media content in the daily production workflow.

**Keywords:** Linked Enterprise Data, Linked Media, Semantic Media Archive

## 1 Introduction

The Linked Open Data (LOD) community project was initiated in 2007 by the W3C [1] and proposes the usage of standards like the Resource Description Framework (RDF) [2] for publishing datasets on the web in order to make them available for interlinking [3]. The number of datasets available, commonly referred to as the Linked Data Cloud<sup>1</sup> [4], is still growing and provides enterprises with the opportunity to interlink enterprise data with background information or to allow for disambiguation of concepts. Enterprises however still hesitate to use Linked Data in their value chain. Based on experiences with industrial partners, the main barriers in the adoption of Linked Data are (i) a rather new technology since accessing data from the Linked Data cloud is still cumbersome; (ii) the lack of complete solutions because Linked Data is still considered read-only and metadata-only whilst enterprise data is highly dynamic and increasingly includes multimedia content and (iii) the need of adapting established enterprise processes when using linked data [5].

---

<sup>1</sup> <http://richard.cyganiak.de/2007/10/lod/>

With this article we propose the integration of large datasets available on the web by following the Linked Data principles as outlined in [6] to enhance closed enterprise content with additional information from the Linked Open Data cloud [7]. This demo uses the Linked Media Framework (LMF<sup>2</sup>) [8], a platform for enterprise information integration. Based on Linked Data as well as Apache Stanbol<sup>3</sup> for content analysis, the LMF shows how to eliminate the entry barriers when using Linked Data in enterprises.

## 2 Semantic Media Archive

The Austrian Public Broadcaster's (ORF – Österreichischer Rundfunk) archive is the central repository for all video and audio material created by the ORF in the last 60 years and contains a vast amount of media content in different formats. The primary objective of the archive is to preserve audio/video content for potential future use and make it accessible to editors. When archiving new content, several archiving tools restricted to expert users are used; FESAD<sup>4</sup> as an example is used to manage video based content. However, for journalists, editors and program planners the archiving division uses a web based tool for federated search and investigation. For now, the work of describing the clips (e.g. annotating the content) is actually carried out solely by members of the archiving division. The users of the search tool currently cannot modify/annotate content in order to improve data quality or search confidence.

The main objective for the ORF is therefore to (i) provide additional information to the end users like editors and journalists, (ii) to allow simple annotation means which are not restricted to the archiving division and (iii) provide/integrate semantic search facilities for improved search results. As an integrated solution we integrated the LMF as Linked Media Server in the Archival Toolset of the ORF. In addition to the existing tools, the LMF provides extended semantic search facilities and also allows for interlinking of archival content with publicly available linked data sources. As shown in Fig. 1, the LMF extends the search tool mARCo by adding itself as an additional data source and by providing means for annotating mARCO search results. The annotations are then subject of future searches in mARCO.

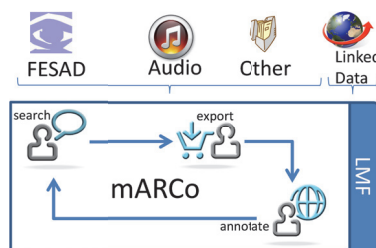


Fig. 1. Semantic Media Archive

<sup>2</sup> <http://code.google.com/p/lmf/>

<sup>3</sup> <http://incubator.apache.org/stanbol/>

<sup>4</sup> FESAD – Video Archival System used by ORF, ARD

## 2.1 Annotating Media Content

When browsing search results, editors or journalists are enabled to annotate the content. With the help of a special annotation plugin, the formerly “read-only” search result page becomes editable by injecting the annotation features into the web page. Parts of the page such as the content description are analyzed by Apache Stanbol<sup>5</sup>. As a result, eligible resource annotations are provided to the user as shown in Fig. 2.



Fig. 2. Annotation and interlinking interface

By selecting a suggestion, the journalist can review the proposal and finally annotate the content. The Linked Media Framework stores the annotation by means of SPARQL Update [10] and also collects the available properties of the referenced resource and thus makes the information immediately available for semantic search.

## 2.2 Semantic Media Search

The search experience can be improved by facilitating the semantic relations of the archived data. By using the semantic concepts of the data which are either production related (e. g. moderator, editor, program etc.) or content related (e.g. persons named or in video, content description, location of the clips content), it is possible to provide a faceted search as shown in Fig. 3, for example to narrow down the search, the user may select one or more facet properties shown in the search interface.



Fig. 3. Search Demonstrator

<sup>5</sup> <http://incubator.apache.org/stanbol>

### 3 DEMO OUTLINE

The Linked Media Framework (LMF) serves as the backend whereas the both clients for search and annotation are lightweight JavaScript implementations using RESTful webservices for the communication with the backend service. The LMF is a service oriented framework which uses semi-structured data representation (RDF) and HTTP URLs as uniform resource identifier to store and identify resources, as recommended for Linked Data [6]. The demo we show at the conference will first show the Semantic Search Component as it is a fundamental part of the LMF and demonstrates the power and flexibility of using Semantic Web technologies for search and retrieval. We will then use a VIE bookmarklet<sup>6</sup> for the annotation of a typical ORF search result page which relies on concepts from DBpedia<sup>7</sup> and an internal SKOS<sup>8</sup> based thesaurus. Accepting proposed annotations with the LMF will immediately influence the search results and optionally add new concepts to an internal company thesaurus. In the production scenario, the LMF will also be tightly connected with the mARCo search facility and therefore will be part of the federated search component.

The LMF integrates/connects the linked data cloud as possible sources for background information and finally enables annotation by storing selected concepts in the (local) Linked Data server by means of SPARQL Update statements. In particular this annotation functionality will be subject of the demonstration given at I-Semantics to first show the where we will preload the LMF with a selection of news articles out of the Austrian Broadcasters Archive. The demonstration will also cover how the news articles are presented to journalists for annotation. Finally, the demonstration of the search interface is also available online at the NewMediaLabs demonstration site<sup>9</sup>.

### 4 CONCLUSION

The potential of Linked Data in general and the Linked Media Framework as a platform for supporting semantic search has been proven in several projects. With this demonstration we aimed to outline its potential for the use in an Enterprise Information Integration scenario where Linked Data technology is used to support users in their daily work and to improve the amount and quality of content annotation. The latter directly leads to an improved search result with respect to precision which is a fundamental requirement in the news domain. Because of the smooth integration in existing processes, the functionality is offered as an optional add-on to the users. The improved search results as well as the provided background information are the inducement for the users to use the offered functionality. In contrast to the increasing number of semantic web case studies<sup>10</sup>, the demonstrated scenario Linked Media Framework allows the publication of structured information as Linked Data and also

---

<sup>6</sup> <http://szabyg.github.com/vie-annotation-bookmarklet/>

<sup>7</sup> <http://dbpedia.org>

<sup>8</sup> <http://www.w3.org/2004/02/skos/>

<sup>9</sup> <http://labs.newmedialab.at/ORF/orf/search/index.html>

<sup>10</sup> <http://www.w3.org/2001/sw/sweo/public/UseCases/>

enables the full read-write management of the published data and in particular enables the full roundtrip of annotations for further usage during search and retrieval.

## 5 ACKNOWLEDGMENTS

The media content enhancement and the semantic search described in this paper were planned and developed in the Austrian research centre "Salzburg NewMediaLab - The Next Generation" (SNML-TNG). The centre is funded by the Austrian Federal Ministry of Economy, Family and Youth (BMWFJ), the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT) and the Province of Salzburg. The demo content is taken from the ORF archive by courtesy of the Austrian Broadcasting Corporation. The development of the LMF has been inspired by the needs & requests of our industrial partners. As a result, the Linked Media Framework currently serves several real-world scenarios.

## 6 References

1. Linking Open Data. 2010. W3C SWEO Community Project. Retrieved from <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
2. RDF: G. Klyne and J. J. Carroll. Resource description framework (RDF): Concepts and abstract syntax. Technical report, W3C, 2 2004
3. Bizer, C., Cyganiak, R., Heath, T. 2007. How to Publish Linked Data on the Web. Retrieved from <http://www4.wiwi.wiwi.fuberlin.de/bizer/pub/LinkedDataTutorial>
4. Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 4(2), 1-22. Elsevier. Retrieved from <http://www.citeulike.org/user/omunoz/article/5008761>
5. Wood, D. 2010. Linking Enterprise Data. ISBN 978-1-4419-7664-2. DOI 10.007/978-1-4419-7665-9
6. Berners-Lee, T. 2006. Linked Data – Design Issues. Retrieved from <http://www.w3.org/DesignIssues/LinkedData.html>
7. Bizer, C., Heath, T., Ayers, D., Raimond, Y. 2007. Interlinking Open Data on the Web (Poster). In 4th European Semantic Web Conference (ESWC2007), pages 802–815.
8. Kurz, T., Schaffert, S., Bürger, T. (2011). LMF – A Framework for Linked Media. In: Workshop for Multimedia on the Web (MMWeb2011).
9. Damjanovic, V., Kurz, T., Westenthaler, R., Behrendt, W., Gruber, A. and Schaffert, S. 2011. Semantic enhancement: The key to massive and heterogeneous data pools. In Proceeding of the 20th International IEEE ERK (Electrotechnical and Computer Science) Conference 2011, Portoroz, Slovenia.
10. Prud'hommeaux, E., & Seaborne, A. 2008. SPARQL Query Language for RDF. W3C working draft. Retrieved from <http://www.w3.org/TR/rdf-sparql-query>