

From Script Idea to TV Rerun - The Idea of Linked Production Data in the Media Value Chain

Harald Sack

Hasso-Plattner-Institute for
IT Systems Engineering
Potsdam, Germany
harald.sack@hpi.de

ABSTRACT

Within the process of the production of a film or tv program a significant amount of metadata is created and - most times - lost again. As a consequence most of this valuable information has to be costly recreated in subsequent steps of media production, distribution, and archival. On the other hand, there is no commonly used metadata exchange format throughout all steps of the media value chain. Furthermore, technical systems and software applications used in the media production process often have proprietary interfaces for data exchange. In the course of the *dwerft* project funded by the German government, metadata exchange through all steps of the media value chain is to be fostered by the application of Linked Data principles. Starting with the idea for a script, metadata from existing systems and applications will be mapped to ontologies to be reused in subsequent production steps. Also for distribution and archival, metadata collected during the production process is a valuable asset to be reused for semantic and exploratory search as well as for intelligent movie recommendation and customized advertising.

Categories and Subject Descriptors

H.4 [Information Systems Applications]:

Keywords

semantic multimedia; data integration; semantic analysis; video analysis; semantic search; recommender systems;

1. Introduction

With the switch from analog to digital technology the entire process of production, distribution, and archival of a film or tv program large amounts of data are created. But, this data is not only restricted to the recorded and processed video and audio information. Moreover each single step within this media value chain creates, administrates, and applies its own metadata that is mandatory for the management of these processes. Unfortunately, a significant amount of this metadata most times is lost again and not available for subsequent steps in the process chain. This is often caused by incompatibilities among software applications, data formats, as well as processing hardware. Companies often want to pursue their own proprietary standards to gain advantages over their competitors and to tie the customer closer to their own products, thus preventing efficient metadata exchange.

As a consequence most of this valuable information has to be costly recreated in subsequent steps of media production, distribution, and archival. Currently, there is no generally accepted and commonly used metadata exchange format that is applied throughout all steps of the media value chain. On the other hand, the market for media production companies has changed with the advent of video on demand and other internet-based marketing platforms. Thus, the market demands a cost and time efficient production setting that follows industry standards, while lost metadata has to be costly reengineered or newly collected.

In the course of the *dwerft* project funded by the German government, metadata exchange through all steps of the media value chain is to be fostered by the application of Linked Data principles. Starting with the idea for a script, metadata from existing systems and applications will be mapped to ontologies to be reused in subsequent production steps. Also for distribution and archival, metadata collected during the production process is a valuable asset to be reused for semantic and exploratory search as well as for intelligent movie recommendation and customized advertising.

2. Linked Production Data

The current Web of Data based on the principles of linked data enables fine granular access to individual data via standardized web technologies (URIs and HTTP) following a simple exchange format (Resource Description Framework, RDF) augmenting the data value by linking it with other (related) datasets [1]. By connecting data with suitable knowledge representations (ontologies) automated data processing and data interpretation is fostered enabling also the deduction of new knowledge from previously implicitly information. By applying the principles of linked data also to metadata being created in the course of the media value chain, efficient and automated reuse of metadata will be enabled as Linked Production Data (LPD).

LPD serves as the core technology connecting the various subtasks in the media value chain, which are organized in media production, media distribution, and media archival. In media production, tasks such as e.g. script writing, production design, costumes, media production, special effects, audio mixing, etc. are subsumed. For each of these tasks, domain experts cooperate to create a suitable terminology describing the task and all associated metadata. This terminology currently is developed with the help of the PoolParty Semantic Suite¹ supported by knowledge engineers, who identify already existing vocabularies, metadata standards, and ontologies to map the created

Copyright is held by the author/owner(s).
WWW '15 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
<http://dx.doi.org/10.1145/2740908.2742917>

¹ <http://www.poolparty.biz/>

terminologies on solid ground. One important metadata standard currently considered for media production tasks within the *dwerft* project is e.g. the Broadcast Metadata Exchange Format (BMF)².

We propose a general adaptable architecture, where technical systems and software applications can map their proprietary interfaces with the open and publicly available Linked Production Data ontologies to enable efficient data exchange.

3. The *dwerft* Project as Application Testbed

The *dwerft* project is publicly funded by the German Government with the goal to establish an industrial setting for media production, distribution, and archival. Besides the development of the Linked Production Data core technology that connects all subtasks within the media value chain, the following compound projects are bundled under *dwerft*:

- Production services are focused on the development of file-based transfer services to fully connect all media production-related tasks.
- Semi-automatic diagnostics and digitization of film material aims to reduce digitization costs dramatically by various automated quality analysis technologies.
- Automated rights acquisition and rights retrieval is focused to open up movie rights to rights distributors.
- Digital Distribution technologies build upon automated content analysis of video data including structural analysis, audio event detection, audio transcription, visual concept detection, and subsequent semantic media analysis.

Semantic media analysis also makes use of LPD in connection with the results achieved by traditional media analysis, such as automated speech recognition or video optical character recognition. In these usually text-based results, named entities are identified and mapped in a disambiguation process to LPD and other linked data resources [2]. Time-based media fragments are identified via scene cut detection and annotated with LPD entities, which can be used to enable semantic and exploratory video search. Semantic video search can be based on entities instead of text keywords to improve retrieval accuracy. Furthermore, content-related entities are applied for faceted filtering of search results to enable efficient navigation.

Exploratory search extends the traditional pinpoint search, where the user exactly knows what she is looking for and how to phrase that search request. But, when searching in an unknown domain, where the user does not know the exact terminology and therefore cannot express exactly her search request, the user will follow a stepwise refinement of the search query taking into account knowledge from previously performed search requests to formulate new requests. In exploratory search, besides the exact search results also documents are of interest that are only related to the original search requests. When searching in an archive of limited expansion, sometimes a search request does not deliver a result, simply because the archive content is limited and does not

exactly contain what the user is looking for. Then, also documents that contain closely related topics and entities might be of interest for the user [3]. Thus, exploratory search enables the discovery of new paths to explore the archive content.

In this context exploratory search makes use of recommendations based on related content, which can be determined with the help of LPD ontologies. Movie recommendations as well as customized advertising taking into account media content as well as personal preferences is based on the same principles as exploratory search.

4. ACKNOWLEDGMENTS

This work has been funded by the German Government, Federal Ministry of Education and Research under the project number 03WKJ4D.

5. REFERENCES

- [1] Heath, T. and Bizer, C. 2011. Linked Data: Evolving the Web into a Global Data. Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. DOI= 10.2200/S00334ED1V01Y201102WBE001
- [2] Steinmetz, N. and Sack, H. 2013. Semantic Multimedia Information Retrieval Based on Contextual Descriptions, In Proc. of 10th Extended Semantic Web Conference (ESWC 2013) - Semantics and Big Data, Lecture Notes in Computer Science LNCS 7882, 283-396.
- [3] Waitelonis, J. and Sack, H. 2012. Towards exploratory video search using linked data, In Multimedia Tools and Applications, Volume 59, Number 2. 645-672. DOI= 10.1007/s11042-011-0733-1.

² <http://bmf.irt.de/eng>