Semantic Web Technologies for
User Generated Content and Digital Distribution
Copyright Management

Roberto García, Universitat de Lleida, Spain

Nick Sincaglia, NueMeta LLC, USA

This talk proposal is about part of the work carried out in the context of the EU project MediaMixer. This project aim was to promote adoption of semantic technologies by the media industry. The proposal was based on a workflow enhanced by semantic technologies, as shown in Figure 1.
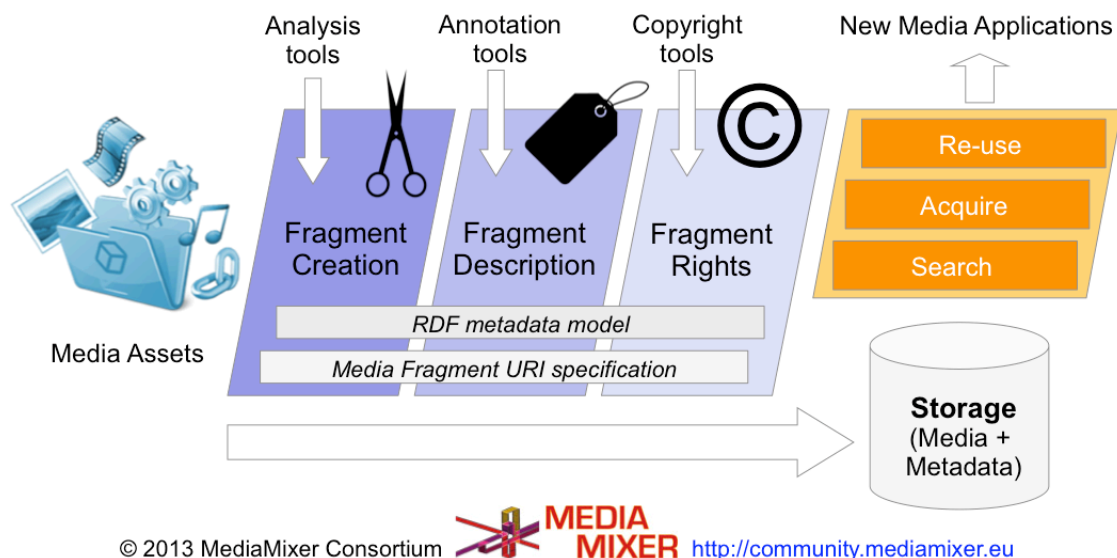

Figure 1. MediaMixer project workflow enhanced with semantic technologies

The first part of the workflow is about automatic fragmentation of media assets to facilitate their reuse and repurposing, using the Media Fragments URI specification. These fragments are then semantically annotated during the second step of the workflow (using automatic, semiautomatic and manual approaches). Annotations are based on semantic data and existing vocabularies and ontologies. Finally, rights information is also attached to fragments. This information is also based on semantic technologies and the Copyright Ontology.

The talk would focus on the last part of the workflow, more specifically on what has been achieved to facilitate copyright management through two real application scenarios.

The first one involved NueMeta LLC, an USA-based technical development and service provider that helps media companies design and develop digital asset management and royalty reporting systems, and Sony DADC, part of the Sony Group and mainly dealing with digital music distribution.

The second scenario is about extending an existing commercial tool by NueMeta LLC to facilitate digital distribution of music based on basic cloud-based infrastructures like Google Spreadsheets and Amazon file storage.

SONY DADC UGC COPYRIGHT MANAGEMENT SCENARIO

UGC is content created by users and shared through platforms like YouTube. Usually, it is the mix of content really generated by the user, like a wedding recording, combined with content owned by others, for instance the couple's favourite song by Adele.

To preserve rights holders rights on copyrighted pieces of content used without permission, UGC platforms provide content identification services so owners can register their content and be warned when it is reused. Though the first reaction might be to just block content reusing copyrighted media without authorisation, UGC platforms have generated a new revenue stream by sharing part of the advertisement revenue generated by content views if it is kept in the platform.

This is becoming an important revenue stream for owners of hits in UGC. However, this requires a big mind change in media management. To foster media remixing and viral reuse of content, content owners should move away from content protection measures like DRM that might prevent their content from being remixed. They should, however, focus on technological measures that facilitate reuse while tracking it and then managing copyright, not just at the end-user level but through the whole value chain of mixes and remixes.

The application for Sony DADC features semantic technologies for copyright management based on copyright law and thus provide the modelling tools to capture copyright statements from sources ranging from digital operations to talent contracts. Moreover, thanks to the reasoning features these technologies provide, these semantic models can be then used to support intelligent decision support at the scale of a media repository and its associated copyright statements.

For instance, it is possible to help deciding the reaction to detecting that a song by Green Day is being reused in YouTube. Should it be monetized and streamed together with advertisements or simply blocked? Semantic technologies allow going beyond just choosing to monetize based on the limited information available at the digital operations stage. The objective is to avoid the legal troubles that might arise from ignoring, at that particular decision point, the fact that the talent contract with Green Day states "*we do not want our creations mixed with war images*". More details are available from a publicly available demo[1], which is not based on real data but illustrates the developed application.

DIYSTRIBUTION DIGITAL DISTRIBUTION TOOL SCENARIO

The DIYstribution tool, developed by NueMeta LLC, is a commercial tool that targets small to medium size content suppliers, and even individuals. The tool constitutes a really affordable option for digitally distributing the media this small players supply. The tool main component is a workflow building and execution application that collects all the metadata about the contents to be distributed from Google Spreadsheets and distributes the actual content among different storage solutions like Google Drive, Amazon S3 or FTP/S services.

The existing workflow execution component is being extended using semantic technologies to make it capable of dealing with a wider range of sources of information about the copyright status of the content being distributed. Currently, the component generates metadata based on DDEX, the music industry standard for digital data exchange and uses it to drive its operations. As a result of the integration of semantic technologies, DDEX data is first mapped to semantic data based on the Copyright Ontology. Then, it becomes easier to integrate it with other sources of information about copyright and to develop more sophisticate decision support scenarios that benefit from the copyright domain knowledge captured by the ontology.

This is currently work in progress and much more will be possible to report and demo by the time of the presentation.

---

[1] MediaMixer UGC Copyright Management Demo, http://rhizomik.net/mediamixer