A Semantic Approach for the Repurposing of Audiovisual Objects

Benjamin Diemert, Marie-Hélène Abel, Claude Moulin
Université de Technologie de Compiègne, France
Email: (benjamin.diemert, marie-helene.abel, claude.moulin)@utc.fr

***Abstract*—In this paper, we address the issue of audiovisual document modeling to promote its repurposing. It requires to identify the various components of the document that can be reused during professional production. We describe a use case which incorporates User Generated Content and considers various exploitation forms and medium. We define a conceptual model and present real applications using it. The main contribution of this paper is the conceptual model which distinguishes between editorial, encoding and sequencing aspect of an audiovisual document.**

***Keywords*-audiovisual content; audiovisual document; repurposing; metaproduction**

I. INTRODUCTION

The main issue at stake in the audiovisual field is the lack of standard formats to support content creation and repurposing. The modeling of the audiovisual document is still an open debate which should not be restricted to the choice of a media wrapper. Indeed, audiovisual documents are not just the digital files delivered to the viewer at the end of the production chain. They are a montage of content picked up in various files and ordered given editorial guidelines to convey a message. A fine modeling of the audiovisual objects is needed to promote a common view on the audiovisual document components and thus foster their repurposing.

A closer look on audiovisual repurposing – also called metaproduction by [1] – reveals various practices ranging from sampling of content fragments, to complete reuse of document's parts via reencoding of video material to fit technical restrictions. Each practice enables the exploitation of a different level of the audiovisual document, i.e. the content assemblage, the editorial structure or the material realisation. Repurposing is thus a desired practice because it reduces production costs and opens up new exploitation forms.

Another significant shift to promote audiovisual repurposing is the ability to integrate User-Generated Contents (UGC) into professional production. Since the widespread of quality camera in mainstream electronic devices, amateurs are now able to create audiovisual content in a glimpse. They can now compete with professionals for the capture of unpredictable events such as natural disasters. However, their lack of training in audiovisual production requires additional efforts to leverage the quality of UGC to professional criteria. This trend enables professionals to diversify their sources of content as well as produce content in a more participative way.

In any case, repurposing opportunities require a shared conceptual model of the audiovisual document components to be seized and exploited. Ontologies as formal representations of a conceptualization can provide such a model. The intent is to provide public conceptual models that can be mapped together and extended to fit more specific use [2]. Hence, an ontology of audiovisual document could be related to content description ontologies but also specialized to handle specific kinds of documents. Information about an audiovisual document created from such an ontology could then be linked to other data on the Semantic Web and thus integrate both media- and domain-dependent descriptions.

In this paper, we present a use case implying UGC integrated into a professional production and show modeling aspects required by it (section II). Then, we explain how our ontology copes with these requirements (section III) and compare it to other related standards and models (section IV). Finally, we show how applications are using our model to provide a solution for UGC repurposing in a similar situation to our use case (section V).

II. USE CASE AND REQUIREMENTS

A. Use Case

Let's consider a public channel interested in the capture of a cultural event such as an opera, a play or a concert. The channel's producers are interested in three kinds of content; *material from the event* itself, *interviews* of the director and performer and *comments from members of the audience*. Producer's concern is to minimize the number of people employed and the equipment used while ensuring the material quality and maximizing its exploitation. Thus, they envisage to split the work and call for amateurs' contributions:

- the capture of the event requires a professional staff with sufficient technical skills and various recording equipments for a long period of time (equipment set up, recording test etc.).
- interviews can be performed by an another team composed of only a journalist and a cameraman.
- comments from the audience members can be shot by the spectators themselves before and after the performance. In order to ensure its quality, they should be assisted and follow guidelines.

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-129-8

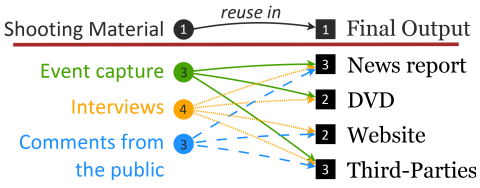

Figure 1. Overview of the material created in the use-case and their exploitation opportunities

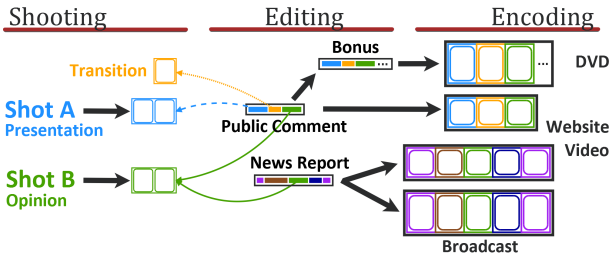

Figure 2. Detailed view on the spectator's comments processing for each exploitation case

Concerning the material's exploitation, Fig.1 depicts 4 possible exploitation cases (in *italic*):

- a *news report* for the channel is created from an assemblage of every kinds of content
- materials from the event are kept for a *DVD* with the interview and best comments from the public as bonus content
- some comments from the public and the interview are reused on the auditorium's *website* for promotion
- any part of the material can be sold to a *third-parties*

All exploitations reuse some shooting material but require to be produced by distinct processes in order to fit the specificities of each distribution medium/channel and meet the expectations of the intended audience. However, the shooting is organised independently from the rest of the production chain. While every process share the resulting material, it is managed differently to produce the final content. This leads to variations in the material encoding quality and the content editorial formatting. For instance, a news report has usually a more intense pace than the bonus content of a DVD and thus different editing of the same material. Moreover, bandwidth constraints are quite different between broadcast and web distribution as between dvd and vhs tape recording.

In our case, the multiplicity of production teams and the diversity of repurposing situations does not admit a simple solution like manual file annotation and sharing. With each team dealing with other's material, there is a need to clear up what result of the process will be reused for each exploitation case. From now on, we focus on the processing of the spectators comments shown Fig. 2:

- Each comment *shooting* is divided in two shots, one for the spectator's presentation and another to express his/her opinion on the performance. These shots are the raw material shared by DVD bonus, promotion website and the news report exploitation case.
- The two shots are *edited* differently according to the editorial structure defined by the professional producers of each exploitation cases.
  - The DVD and the promotion website share also parts of the editing line. Both assemble the shots together with a transition in between. A selection is directly published on the website while another editing with a compilation of the best ones is created for the DVD bonus content.
  - For the news report, only the opinion shot is integrated into the report structure along with jingles and comments from a journalist.
- Each final edit is then *encoded* specifically to fit the distribution constraints, the highest quality for DVD and channel broadcast then a lower quality for the website distribution. Note that the news report is distributed both on the public channel and on the channel website, resulting in two different encoding of the same edit.

B. Modeling requirements

The details of this use case specifies more than a simple reuse of material. It specifies the kind of processing that support repurposing and which defines thus the modeling requirements for the audiovisual document:

- the *reencoding* of edited material to fit the technical parameters proper to each distribution medium/channel (news report distributed by channel broadcasting and internet).
- the reuse of shooting materials in two distinct editorial structure (*resequencing* of the opinion shot in the website and news report editing).
- the reuse of a part of an editorial structure into another editorial structure (*repurposing* of the public comment editing into the DVD bonus editing).

In the next section, we define a conceptual structure which represent the audiovisual document. The principle of our modeling is to identify distinct components of the audiovisual document in order to handle them separetely and thus enables these three kinds of reuse.

III. Audiovisual Object Model

The purpose of this section is to detail the modeling of an audiovisual document. Firstly, we clarify the distinction between video material and content sequence (subsection III-A), then present an example extracted from the use case which illustrates reencoding and resequencing (subsection III-B). Secondly, we distinguish between content montage and editorial structure (subsection III-C) then show an example which depicts repurposing (subsection III-D).

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-129-8

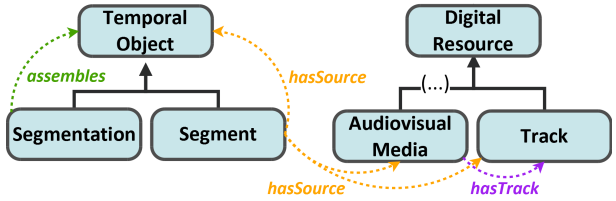

Figure 3. Concepts representing digital file, audiovisual content and their relations

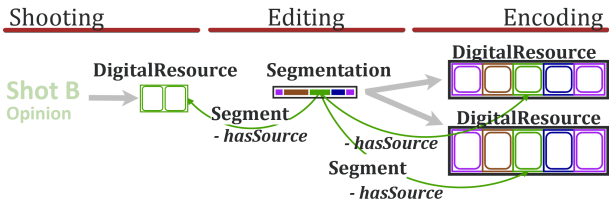

Figure 4. Modeling of the opinion shot used for the creation of a News Report

A. Resource versus Content

The challenge of audiovisual object modeling is to distinguish its various components and relate them. The audiovisual object is all together a recording, a content and a document. The recording is a digital file stored on a medium. The content refers to the viewing or the playing of this recording, i.e. what can be actually perceived by a human or sensed by a machine. The document is a communicative object created and structured in order to convey a message to others.

In this first part, we will focus on the recording and playing part of the audiovisual object. The recording is created from the encoding of a video flow with a compression algorihtm such as MPEG-2 or H.264 and encaspulated in a digital file according to a wrapper format such as AVI or MKV. The audiovisual content flow is reconstructed from the decoding of a series of bits. One important feature of this relation is that content can be recreated from various recording – with different encodings for instance.

From this first description of the audiovisual object, we have defined a distinction already highlighted by [3] between its recording form (*DigitalResource*) and its playing form (*TemporalContent*). Fig. 3 depicts the main relations between these concepts and their specializations:

- A **DigitalResource** is a sequence of bits with its own address (*accessPath* attribute), like a digital file or a portion of a digital file. We distinguish between two kinds of DigitalResource:
  - a **MediaWrapper** is a DigitalResource which encapsulates any kind of media resources (audiovisual, picture, text etc.) according to a given format (*wrapperFormat* attribute). A MediaWrapper details the name of the encoding algorithm used to create it (*encodingMethod* attribute which in the case of video use the FourCC identifier. FourCC provides a four character code identifier as well as a short description for 331 video codecs. A MediaWrapper also specifies the main encoding parameters such as *samplingRate* (frequence of value picking) and *bitResolution* (value encoding precision). We specialize this concept for resource holding audiovisual content with the concept of *AudiovisualMedia*.
  - a **Track** is a portion of a digital file holding a series of content which may be unwounded to create temporal content, like an audiovisual, audio or subtitle track. Track shares the same attribute than MediaWrapper to define its encoding method and parameters.
- A **TemporalObject** is an object which content is unwinded and revealed progressively during a fixed period of time (*objectDuration* attribute). We distinguish two kinds of TemporalObject:
  - a **Segment** is a temporal selection on a flow. It uses the timecode system from the source as a reference to define its starting position in the flow (*timeCodeIn* attribute) and its duration to define its ending position. The source of a Segment (relation *hasSource*) designates the object on which the timecode selection is made. It can be one or more TemporalObject, MediaWrapper or Track.
  - a **Segmentation** is an ordered collection of TemporalObject. Each TemporalObject is related to the Segmentation by an *assembles* relation. Segmentations can be used to represent the temporal content montage or a particular content indexing.

The goal of this modeling is to keep record of the relation between content and the ressources which provides the video material. We explain now how to use these concepts to model the reencoding and resequencing of audiovisual material.

B. Reencoding and Resequencing

With the concepts of DigitalResource and TemporalObject, we can represent the three stages of our use case like shown Fig.4:

- The shooting stage creates DigitalResources, such as the audience's member opinion shot.
- The editing stage creates a Segmentation which represent a simpler version of the actual editing line. The Segmentation is composed of Segments pointing to the DigitalResource used as raw material – thanks to the hasSource relation. In our scheme, the hasSource

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-129-8

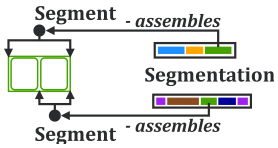

Figure 5. Resequencing of the same DigitalResource implies two different Segments here included in two Segmentations

attribute linking the opinion shot DigitalResource to its Segment is represented by a thin green arrow line.

- The Encoding stage creates two DigitalResources as outputs. One Segment is enough to represent the content sequence corresponding to the opinion shot and encoded differently in each DigitalResource. Indeed, the Segment has the same beginning and duration but points out to distinct DigitalResource through distinct hasSource relations.

This example shows that the *hasSource* relation between a Segment and a DigitalResource allows a multiplicity of technical variations (encoding method, wrapper format) and copies to be linked to the same content. This linking enables thus to deal with reencoding of the same content.

In the same example, if some part of the opinion shot needed to be cut away for the news report, then this selection would be represented by another Segment with another beginning and duration. Fig.5 shows an overview of this case where the news report Segmentation would rather integrate this Segment. In addition, the selection could be directly made on the spectator's comment Segmentation as it is also a temporal object with a duration. Thus, the Segment concept allows us to deal with the resequencing of content into various Segmentations.

C. Content versus Editorial Structure

An audiovisual object is not only a content recorded on a medium, but also a document created by humans to convey a message to other humans. Thereby, the audiovisual object is related to an intended message defined during pre-production, an actual content realized during production and a perceived message interpreted during the viewing. Usually, a document refers to a genre which prescribes a pattern to structure the message. The genre pattern aims at easing the message transmission by providing a common reference to the document's creators and viewers. For instance, everybody expects an interview to be composed by a series of questions and answers.

We distinguish between the document which refers to a genre and the editorial objects which form its actual structure. A common view is to consider the document as a production result while the editorial objects represent the creation units. The document is created one editorial object at a time, but it is ordered, sold and distributed as a whole. At pre-production stage, the editorial objects hold intentions

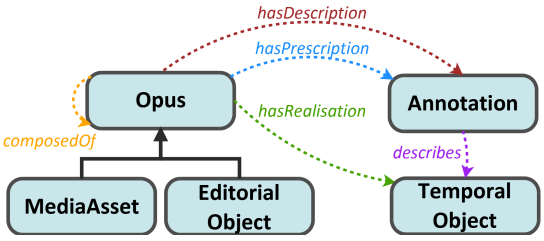

Figure 6. Concepts representing content, editorial structure, annotation and their relations

on the content to be produced. At production stage, materials are created or acquired to express at best these intentions. At the editing stage, contents montage can reveal several particular points of view on the way to realize the original intentions. At post-production, the editing is finalized and the document can be packaged for distribution.

From these details on the fabrication of the audiovisual object, we have defined a distinction between an editorial perspective and the assemblage of video material. This distinction enables us to separate the editorial structure (*EditorialObject*) of audiovisual document (*MediaAsset*) from its potential realisations (TemporalObject defined previously). In addition, we can describe each of these elements with specific *Annotation* such as script extract, dialogue, signal analysis etc. Our Annotation concept is generic and extensible as defined in [4]. We do not defined it further in order to focus on audiovisual objects modeling, even if we are well aware that annotations are needed to describe the audiovisual object components. We just point out that different kinds of annotation can be used and explain where they should be attached. Fig.6 depicts the main relations between these concepts:

- a **MediaAsset** is a document intended to be published. Its structure varies for each specialization of media asset and defines thus a new genre or format. For instance, an interview is composed of questions and answers.
- an **EditorialObject** is a document's fragment which composes the editorial structure of the MediaAsset. EditorialObject's composition can also be structured by editorial rules. For instance, a shot can be divided in subshots to detail complex camera movements like travelling followed by a panning. For audiovisual documents, we specify generic and basic kinds of editorial objects:
  - **Shot** is a series of uncut frames, which means a group of pictures recorded between two pushes on the record button (on and off). Shot can be composed of SubShot or FilmedElement.
  - **SubShot** is a part of a Shot with a continuous camera movement or a constant framing.
  - **FilmedElement** is something that appears in the

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-129-8

frame, usually a main part like an actor, an object etc. rather than the background.

MediaAsset and EditorialObject can be merged into the more generic concept of **Opus**. An Opus is a repurposable piece of work with an editorial consistency. Its purpose is to enrich the modeling of the piece of work by binding it to other concepts:

- a prescriptive annotation which reflects the editorial intentions to realize (relation *hasPrescription* with Annotation as range). Such an annotation can integrate script extract or dialogue.
- a portion of video material which has either been the result of an original creation or a selection of existing content (relation *hasRealisation* with TemporalObject as range).
- other Opus which contribute to define an editorial structure (*composedOf* relation).
- descriptive annotations which reflect the editorial choices made during production (relation *hasDescription* with Annotation as range). Such annotations can be considered as an updated and extended version of the prescriptive annotations. They can integrate extracts from the final script, actual dialogue but also conceptual indexing from a controlled vocabulary or an ontology.

Note that there is another kind of Annotation which is directly connected to a content sequence (relation *describes* with TemporalObject as range). These Annotation can be composed of video material analysis which are specific to the content and not to the editorial choices made.

D. Repurposing

The goal of repurposing is to support the reuse of editorial objects inside various documents. As an example, we explain how the spectator's comment can be repurposed. The whole comment is modeled as a MediaAsset, while the presentation and opinion shots are EditorialObjects. After the shooting, two similar selections (Segment) are made on the best opinion shot (DigitalResource). One very short for the news report, another much larger for the promotion website and the DVD bonus content. In this case, the *hasRealisation* relation between EditorialObject and Segment allows the two selections to be related to the same editorial object. As a consequence, each content sequences is easier to find from the other and both benefit from the annotation of the editorial object.

After the montage of the spectator's comments, the editor of the DVD bonus content wants to reuse the montage of the best comments. In Fig.7, we depict our modeling of this example. The bonus content is modeled as another document (MediaAsset) which reuses some existing comments (MediaAsset). Here, the repurposing is made without changes in the editing of the reused documents. The *composedOf* relation between two MediaAssets represents the hierarchical integration of one in the other while the montage is

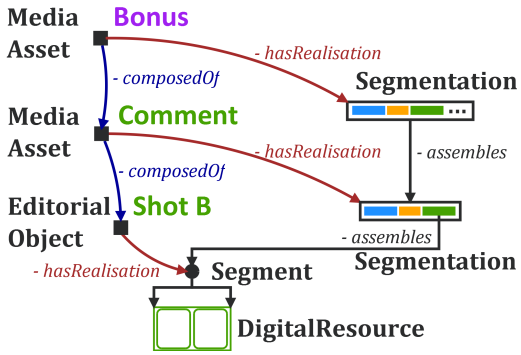

Figure 7. Repurposing of public comments in DVD bonus content

modeled by a Segmentation. In this case, the bonus montage (Segmentation) is made by an ordered assemblage of the comment's montage (Segmentation) through the *assembles* relation.

IV. Related Works

Despite a full conversion of the prominent MPEG-7 standard into a OWL Full ontology achieved by [5], the syntactic and semantic ambiguity of MPEG-7 demonstrated by [6] remains a real concern for data integration. The COMM ontology has clarified how formal semantics could be added to the MPEG-7 using patterns from the DOLCE foundational ontology [2]. COMM also contributed to highlight the need for separation and interrelation between low-level signal features, content sequence and annotation. Compared to COMM, our model defines an additional representation level to cope with editorial composition. With this additional level, our model enables to manage video material, content sequence and editorial composition indenpendently.

V. Applications

The audiovisual object model presented in the previous section has been developped in the course of the MediaMap project [7]. MediaMap is a Celtic project which aims at innovating in the area of audiovisual content production, in particular in the niche of UGC. In this project, we have formalized our model into an OWL-DL ontology so it is used in the project's applications: a shooting assistant for UGC production developped by SkemA and a search interface dedicated to audiovisual professionals developped by Exalead. SkemA is specialized in the developpment of Web and Mobile video platforms and Exalead a solution provider for entreprise and web search. Both applications use our ontology for representing an actual situation similar to the use case described in section II. That situation concerned the Tannhaüser opera, composed by Richard Wagner and directed by Jan Fabre.

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-129-8

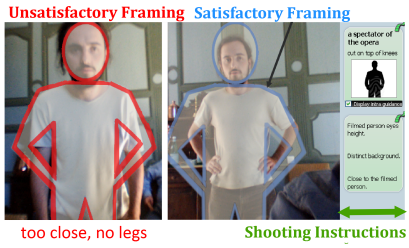

Figure 8. Guidance during shooting provides a summary of the instructions and a real-time framing indicator

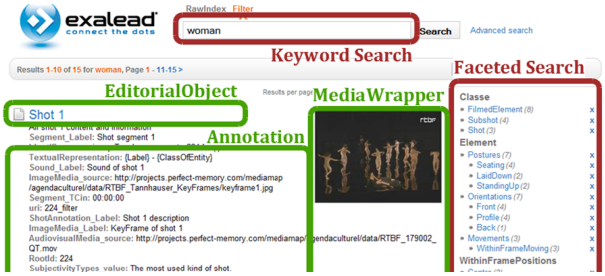

Figure 9. Search interface with keyword and faceted search on audiovisual objects

A. UGC Shooting Assistant

The "CameraMate" shooting assistant is a software embedded into an ad-hoc prototype camera developped by Vitec Multimedia. Vitec Multimedia is a hardware provider specialized in digital video equipments. It is designed with a form-like interface to manage the shooting workflow for the amateur cameraman. Guidance are given before and during shooting according to predefined editorial recommendations defined using an audiovisual scripting vocabulary. This high-level vocabulary is related to low-level signal analysis algorithms which provide real-time indicators of the relevancy of the shooting according to the recommendations – see Fig.8.

The output of the CameraMate is conform to the examples detailed in the previous section. For instance, each shot is modeled as an EditorialObject with the shooting recommendations as prescriptive Annotations. The video material captured by the camera is a MediaWrapper. Finally, when the mission is done the cameraman can send the result to the professional producer which ordered the shooting.

B. Audiovisual Search for Producer

Once the UGC is retrieved by the professional producers it is stored in a semantic repository developped by Memnon. Memnon is a service and solution provider for media digitization and archiving. Exalead is in charge of indexing the audiovisual objects and provides then a dedicated-version of its search solution, as shown Fig.9. Professional can use the interface to retrieve any kind of audiovisual objects described before and thus enables repurposing. The results are presented as attached to an editorial object (MediaAsset or EditorialObject). Annotations provide general and audiovisual descriptions of the editorial object. MediaWrappers provide video and encoding parameters. The interface has two distinct search methods:

- from keywords indexing full-text annotation of audiovisual objects, like general description.
- from existing attributes and their values (faceted search) which provide a mechanism to filter the result set. This feature offers progressive filtering possibilities enabling a conceptual navigation in the result set thanks to our Annotation model.

VI. Conclusion

In this article we have defined a conceptual model of the audiovisual document. Our work contributes to identify and clarify the definition of all the objects composing an audiovisual document. We present a use case involving professional and UGC productions to demonstrate its expressiveness. Finally, we present applications using our model to provide shooting assistance and enhanced audiovisual search. We are currently working on the evaluation of our model in the more general context of collaborative audiovisual production including web 2.0 technologies. For that purpose, we consider essential the use of an organizational memory. This can be seen as a platform with different services fostering the exchange of knowledge, information and audiovisual resources [8] inside a company.

References

[1] J. V. Ossenbruggen, F. Nack, and L. Hardman, "That obscure object of desire: multimedia metadata on the Web, Part-1," *Multimedia, IEEE*, pp. 38–48, 2004.

[2] R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura, "COMM: designing a well-founded multimedia ontology for the web," *The Semantic Web*, pp. 30–43, 2007.

[3] P. Morizet-Mahoudeaux and B. Bachimont, "Indexing and Mining Audiovisual Data," *Lecture Notes in Computer Science*, vol. 3430, no. 5, pp. 34–58, 2005.

[4] L. Hardman, v. Obrenovi, F. Nack, B. Kerhervé, and K. Piersol, "Canonical processes of semantically annotated media production," *Multimedia Systems*, pp. 327–340, 2008.

[5] R. Garcia and O. Celma, "Semantic Integration and Retrieval of Multimedia Metadata," in *SemAnnot 2005*, Ireland, 2005.

[6] F. Nack, J. V. Ossenbruggen, and L. Hardman, "That obscure object of desire: multimedia metadata on the Web, part 2," *Multimedia, IEEE*, vol. 12, no. 1, pp. 54–63, Jan. 2005.

[7] "Mediamap project," Web site: http://www.mediamapproject.org/ [Last accessed: February 2011].

[8] M.-H. Abel and A. Leblanc, "Knowledge Sharing via the E-MEMORAe2.0 Platform," in *ICICKM 2009*, Montreal, Canada, 2009, pp. 10–19.

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-129-8