Integration of Existing Multimedia Metadata Formats and Metadata Standards in the M3O

Daniel Eißing, Ansgar Scherp, and Carsten Saathoff

WeST, University of Koblenz-Landau, Germany
{eissing,scherp,saathoff}@uni-koblenz.de
http://west.uni-koblenz.de

Abstract. With the Multimedia Metadata Ontology (M3O), we have developed a sophisticated model for representing among others the annotation, decomposition, and provenance of multimedia metadata. The goal of the M3O is to integrate existing metadata standards and metadata formats rather than replacing them. To this end, the M3O provides a scaffold needed to represent multimedia metadata. Being an abstract model for multimedia metadata, it is not straightforward how to use and specialize the M3O for concrete application requirements and existing metadata formats and metadata standards.

In this paper, we present a step-by-step alignment method describing how to integrate and leverage existing multimedia metadata standards and metadata formats in the M3O in order to use them in a concrete application. We demonstrate our approach by integrating three existing metadata models: the Core Ontology on Multimedia (COMM), which is a formalization of the multimedia metadata standard MPEG-7, the Ontology for Media Resource of the W3C, and the widely known industry standard EXIF for image metadata.

1 Introduction

A densely populated jungle with a myriad of partially competing species of different colors and size—this might be a good characterization of today's world of multimedia metadata formats and metadata standards. Looking at the existing metadata models like [1,2,3,4,5] and metadata standards such as [6,7,8,9,10], we find it hard to decide which of them to use in a complex multimedia application. They focus on different media types, are very generic or designed for a specific application domain, and overlap in the functionality provided.

However, building a complex multimedia application often requires using several of these standards *together*, e.g., when different tools have to be integrated along the media production process [11]. The integration among tools requires interoperability of different metadata standards, which is a requirement that is not sufficiently satisfied by existing formats and standards. With XMP [7], there exists an important initiative to enable interoperability along the production process of images. Nevertheless, this work is limited with respect to the functionality provided and focuses on the media type image only [12]. Overall, the

T. Declerck et al. (Eds.): SAMT 2010, LNCS 6725, pp. 48–63, 2011.
© Springer-Verlag Berlin Heidelberg 2011

XMP initiative is an important step but more is required to facilitate multimedia metadata interoperability along the media production process.

To solve this problem, we have developed the Multimedia Metadata Ontology (M3O) [12]. The M3O is a sophisticated model for representing among others the annotation, decomposition, and provenance of multimedia content and multimedia metadata. The goal of the M3O is to provide a framework for the integration of existing metadata formats and metadata standards rather than replacing them. The M3O bases on a foundational ontology and by this inherits its rich axiomatization. It follows a pattern-based ontology design approach, which allows the M3O to arrange the different functionalities for representing multimedia metadata into smaller, modular, and reusable units.

However, the M3O was designed as an abstract model providing a scaffold for representing arbitrary multimedia metadata. As such, the integration of existing standards is not straightforward, and we are confronted with a gap between the formal model and its application in concrete domains. In this paper, we fill this gap and present a step-by-step alignment method describing how to integrate existing formats and standards for multimedia metadata and the M3O. We describe the tasks that have to be performed for this integration and apply the integrated ontology to a concrete modeling task. We demonstrate this integration at the example of the Core Ontology on Multimedia (COMM) [2], which is a formalization of the multimedia metadata standard MPEG-7 [9], the recently released Ontology for Media Resource [13] of the W3C, and the widely known and adopted industry standard EXIF [6] for image metadata.

2 Introduction to the Multimedia Metadata Ontology

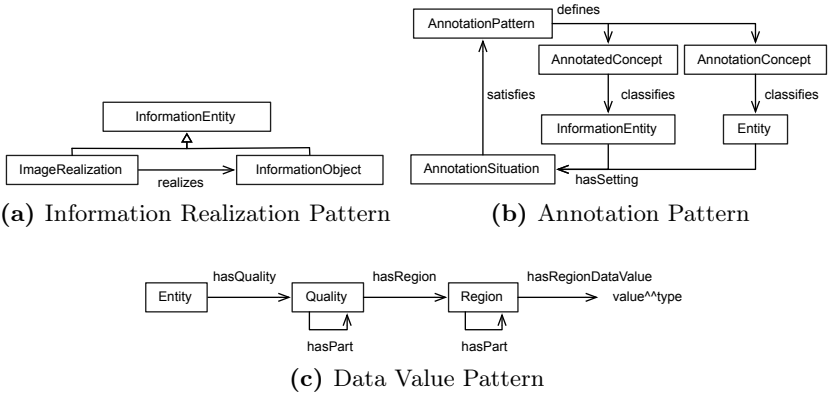
The Multimedia Metadata Ontology (M3O) [12] provides a generic modeling framework to integrate existing multimedia metadata formats and metadata standards. The M3O is modeled as a highly axiomatized core ontology basing on the foundational ontology DOLCE+DnS Ultralight (DUL) [14]. DUL provides a philosophically grounded conceptualization of the most generic concepts such as objects, events, and information. The axiomatization is formulated in Description Logics [15].

The M3O follows a pattern-based approach to ontology design. Each pattern is focused on modeling a specific and clearly identified aspect of the domain. From an analysis of existing multimedia metadata formats and metadata standards [12], we have identified six patterns required to express the metadata for multimedia content. These patterns model the basic structural elements of existing metadata models and are the Decomposition Pattern, Annotation Pattern, Information Realization Pattern, Data Value Pattern for representing complex values, Collection Pattern, and Provenance Pattern. Basing a model like the M3O on ontology design patterns ensures a high degree of modularity and extensibility, while at the same time a high degree of axiomatization and thus semantic precision is retained. In order to realize a specific multimedia metadata format or metadata standard in M3O, these patterns need to be specialized. In

the following, we discuss three patterns of the M3O in more detail, namely the *Information Realization Pattern*, *Annotation Pattern*, and *Data Value Pattern*, which we will mainly refer to in the upcoming sections.

Information Realization. The information realization pattern in Figure 1a models the distinction between information objects and information realizations [14]. Consider a digital image that is stored on the hard disk in several formats and resolutions. An information object represents the image as an abstract *concept* or *idea*, namely the information object of an image. Different files may *realize* this same abstract idea. As shown in Figure 1a, the pattern consists of the InformationRealization that is connected to the InformationObject by the realizes relation. Both are subconcepts of InformationEntity, which allows treating information in a general sense as we will see in the Annotation Pattern.

Annotation Pattern. Annotations are understood in the M3O as the attachment of metadata to an information entity. Metadata comes in various forms such as low-level descriptors obtained by automatic methods, non-visual information covering authorship or technical details, or semantic annotation aiming at a formal and machine-understandable representation of the contents. Our Annotation Pattern models the basic structure that underlies all types of annotation. This allows for assigning arbitrary annotations to information entities while providing the means for modeling provenance and context. In Figure 1b, we see that an annotation is not modeled as a direct relationship between some media item and an annotation. It is defined by a more complex structure, which is inherited by the Descriptions and Situations Pattern of DUL. Basically, a Descriptions and Situations Pattern is two-layered. The *Description* defines the structure, in this case of an annotation, which contains some entity that is annotated and some entity that represents the metadata. The *Situation* contains the concrete entities for which we want to express the annotation. The pattern allows us to add further concepts and entities into the *context* of an annotation, e.g.,



(a) Information Realization Pattern **(b)** Annotation Pattern

(c) Data Value Pattern

Fig. 1. Three Example Patterns of the Multimedia Metadata Ontology (M3O)

expressing provenance or confidence information. On the top half, we see that the AnnotationPattern defines an AnnotatedConcept and an AnnotationConcept. The AnnotatedConcept classifies an InformationEntity and thus expresses that the information entity is the subject of the annotation. The AnnotationConcept classifies some Entity, which identifies this entity as the annotation or metadata. The entity can be some complex data value, e.g., representing some low-level features represented using the Data Value Pattern, but also some concept located in a domain ontology such as DBpedia[1]. All the entities have as setting the AnnotationSituation, which satisfies the AnnotationPattern.

Data Value Pattern. In ontologies we mainly use abstract concepts and clearly identifiable individuals to represent data. However, we also need the means to represent concrete data values such as strings and numerical values. In DUL, there exists the concept Quality in order to represent attributes of an Entity, i.e., attributes that only exist together with the Entity. Regions are used to represent the values of an Quality and the data space they come from. The Data Value Pattern depicted in Figure 1c assigns a concrete data value to an attribute of that entity. The attribute is represented by the concept Quality and is connected to the Entity by the hasQuality property. The Quality is connected to a Region by the hasRegion relation. The Region models the data space the value comes from. We attach the concrete value to the Region using the relation hasRegionDataValue. The data value is encoded using typed literals, i.e., the datatype can be specified using XML Schema Datatypes [16].

3 Alignment Method

This section introduces our method for aligning multimedia standards and multimedia formats with the M3O. The method has been derived from our experiences applying and specializing the M3O for three existing multimedia formats and standards, namely COMM [2], the Ontology for Media Resource [13], and EXIF [6].

In contrast to automatic, adaptive, or machine learning approaches for ontology alignment [17,18,19], we conduct a pure manual alignment, as only a manual alignment ensures the high quality of the integration and minimizes ambiguities and imprecise matching. We consider the time and effort for manual alignment manageable, as we assume that each metadata format or standard has to be aligned only once and that updates to the integrated formats or standards will be infrequent and mostly incremental.

For the alignment, we propose an iterative four-step alignment method that helps ontology engineers to integrate existing metadata standards and metadata formats. In each iteration, we consecutively evolve the alignment of the format or standard with the M3O. Following an iterative approach, we are able to identify, analyze, and flexibly react to problems and challenges encountered during previous iterations of the alignment.

[1] `http://dbpedia.org/`

Each iteration consists of four steps. The first step targets the understanding of the format or standard to be integrated. The second step reorders concepts into coherent groups. The third step maps concepts and structure with the M3O. The fourth step proves and documents the validity of the alignment, and finalizes the iteration.

To introduce our alignment method, we proceed as follows: For each step, we first outline the goals and provide a brief summarization. Subsequently, we describe the core tasks to be performed within the step, and provide concrete examples that show its relevance and application for concrete ontologies.

3.1 Step 1: Understanding

Summary. A precise understanding of the metadata format or standard to be integrated is an import prerequisite for aligning it with the M3O. Consequently, the first step of alignment is an in-depth analysis of the structure and core concepts of the model at hand. While this advise may seem obvious, this is a task easily underestimated and problems neglected at an early stage can cause time-consuming problems along the integration process. Additional documentation, if available, will help to (re-)produce the overall structure not explicitly expressed in the formal specification.

Detailed Description and Examples. In general, we have found three distinct modeling approaches to be very common for multimedia metadata formats and metadata standards.

Pattern-based. Pattern-based ontologies, e.g., COMM, provide a high degree of axiomatization and structure in a formal and precise way. Through our analysis, we understand the patterns used and the functionality they provide. This allows us to compare the patterns of the ontology to be integrated with those provided by the M3O.

Predicate-centric. In a predicate-centric approach, as followed e.g. by the Ontology for Media Resource, the ontology is mainly specified through a set of properties. Such a model offers very little structure in a machine readable format, e.g., in terms of conceptual relations between properties. However, by analyzing the documentation, we infer additional information about intended groupings of properties and the structural composition of the format or standard to be integrated.

Legacy Models. Other formats and standards have not yet been semantified at all. By analyzing the concepts and relations expressed in the specification of the format, we decide how the core concepts can be expressed in a formal and precise way using the M3O.

Ambiguities that are found during the initial analysis are discussed at this point. It is not our intention to revise all modeling decisions made for the ontology to be integrated. However, we consider the alignment a good opportunity to correct some of the *bad smells* [20] discovered. Once we have reached a sufficient understanding of the format or standard to be integrated, we proceed with the grouping step.

3.2 Step 2: Grouping

Summary. Ontologies should provide structural information on the relation and groupings of concepts it defines. However, although many formats or standards provide this information in their documentation, the information is sometimes lost when the models are transformed to an ontology. By using the original specifications and documentations, we are able to preserve and recreate this information grouping, and provide them through formal specification in the aligned ontology.

Detailed Description and Examples. In principle, we distinguish three forms of available grouping information:

Explicit Grouping. Pattern-based models provide an explicit grouping of concepts into coherent patterns, often accompanied by a rich axiomatization on how they relate. As an example, the definition of a color histogram annotation in COMM specifies a ColorQuantizationComponentDescriptorParameter that groups the concepts ColorComponent and NumberOfBinsPerComponent

Implicit Grouping. For other metadata models grouping information may not be explicitly represented. This is often the case with predicate-centric approaches, e.g., the Ontology on Media Resource. In these cases, we refer to the textual documentation in order to (re-)construct the implicit groupings of the properties or classes. As an example, the documentation of the Ontology on Media Resource offers a textual description on the grouping of its properties, e.g., in terms of identification or creation. However, this information is not accessible in the RDF representation as proposed by the W3C. By defining the appropriate axioms, we have appended the implicit grouping information in a formal and explicit way, e.g., by stating that an IdentiÞcationAnnotation hasPart some TitleAnnotation, LanguageAnnotation, and LocatorAnnotation

Recovery of Groupings. In other cases grouping information is lost when transferring multimedia formats or standards to RDF. For example the EXIF metadata standard provides textual descriptions about groupings, e.g., in terms of pixel composition and geo location. However, this distinction got lost in the adaption to an RDF schema [21]. For the alignment with the M3O, we have reconstructed the grouping information and appended it to the model in a formal and explicit way.

Once we have provided all relevant grouping relations through a formal specification, we continue with the mapping step.

3.3 Step 3: Mapping

Summary. This step achieves the mapping of the ontology's concepts and structure to the scaffold provided by the M3O. The goal of this step is to create a working ontology, which, after validation, can be published or used as basis for further iterations.

Detailed Description and Examples. For the alignment we follow a sequence of the following three steps:

1. **Mapping of Concepts.** If some superclass of the concept to be aligned is present in both ontologies, direct mapping of concepts is feasible. This is mainly the case for ontologies that share the same foundation, e.g., COMM and the M3O, which both base on the DUL foundational ontology. All axioms of the aligned concepts are preserved as long as they are applicable through the M3O. If a concept is not applicable in the M3O, we align all dependent subclasses and references to the nearest matching M3O concept. As an example, the COMM DigitalData concept, which is a subclass of the DUL InformationObject, was removed during the alignment. The dangling dependencies and references have been resolved by subclassing or referencing the InformationObjectinstead.
2. **Structural Mapping.** For structural mapping, we consider the functionality of the pattern or structure to be mapped. If a pattern or structure offers the same or an equal functionality than a pattern of the M3O, we can replace the pattern. By adapting the M3O pattern, we are often able to express the same functionality using a more generic approach. As an example, COMM proposes the Digital Data Pattern to express data values in a digital domain. A similar functionality is provided by the M3O Data Value Pattern, which expresses data values through the generic concepts of Quality and Region The COMM Digital Data Pattern can be considered a special case of expressing data values and therefor has been replaced using the M3O Data Value Pattern instead.

   In the same manner, we simplify the structural composition of the existing model by merging multiple concepts and patterns that offer the same or an equal functionality. As an example, COMM defines three annotation patterns. Each deals with a different aspect of multimedia annotation, although they vary only slightly in their structural composition. We have aligned those patterns by adapting the M3O Annotation Pattern. The domain specific concepts that result from the separation into three distinct patterns have been preserved by subclassing the corresponding concepts of the M3O Annotation Pattern. This simplifies the structure of the model, while also preserving the original functionality.
3. **Removing Unnecessary Concepts.** We finalize the mapping step by cleaning up unused dependencies from the ontology files. Concepts that either have no further relevance for the target context or are sufficiently covered by the M3O are removed at this point. An example, the COMM AnnotatedMediaRoleoffers an equal functionality as the M3O Annotated-InformationRealizationConceptWe therefore have removed COMM's AnnoatedMediaRole and replaced any formal relation that involves the concept.

3.4 Step 4: Validation and Documentation

In each iteration of the alignment process, we need to check the consistency of the resulting ontology. This can be done by using a reasoner like Fact++[2] or Pellet[3]. Any problem encountered during the alignment can be resolved by reiterating the four steps of the alignment method. After proving the consistency of the resulting ontology, we finalize the process by documenting all major decisions and adjustments made during the alignment.

3.5 Summary

In this section, we have proposed a four-step method for aligning multimedia metadata formats and multimedia metadata standards with the M3O. In the following Sections 4-6, we demonstrate the alignment of three existing formats and standards by applying our method. They are the Core Ontology on Multimedia, the Ontology for Media Resource, and the EXIF metadata standard.

4 Example 1: Core Ontology on Multimedia (COMM)

The Core Ontology on Multimedia (COMM) [2] is a formal specification of the MPEG-7 metadata standard [9]. In contrast to other approaches to modeling MPEG-7 as an ontology COMM is not designed as a one-to-one mapping, but provides a set of patterns that cover the core and repetitive building blocks of MPEG-7. The central challenge of the alignment of COMM and M3O is understanding the patterns of COMM and mapping them to the scaffold provided by the M3O. This section describes the experiences and challenges of aligning COMM and the M3O, using the four-step alignment method proposed above.

4.1 Application of the Alignment Method

Understanding. COMM follows a pattern-based approach and builds on the DUL foundational ontology. Some of the core patterns, i.e., the Descriptions and Situations Pattern, are shared between COMM and the M3O. Others, e.g., the Digital Data Pattern, form major structural differences.

COMM defines five structural patterns, namely the Content Annotation Pattern, Media Annotation Pattern, and Semantic Annotation Pattern for media annotation, the Decomposition Pattern for media (de-)composition, and the Digital Data Pattern, which expresses annotations in a digital domain. Domain specific knowledge is separated from the core concepts and defined in separate ontologies, e.g., concepts concerning annotation of visual entities are defined in the *visual ontology*.

Some ambiguities that were found in the initial analysis have been resolved at this point. As an example, COMM specifies concepts such as NumberOfBinsPerComponent that are specialization of both Parameter and Region. While this may

[2] http://owl.man.ac.uk/factplusplus/
[3] http://clarkparsia.com/pellet/

not be syntactically incorrect, it violates the DnS pattern of DUL. In the DnS pattern, a Parameter parametrizes a Region. Thus, these two concepts should not have common sub-concepts. To solve this problem, we have removed the superclass relations to the Parameter concept and introduced a parametrizes relation. For example, COMM specified a ColorComponent and NumberOfBinsPerComponent, which are subclasses of both the ColorQuantizationComponentDescriptorParameter and the Region concept. We have removed the superclass relation from the ColorComponent and NumberOfBinsPerComponent to the ColorQuantizationComponentDescriptorParameter, which instead now parametrizes these concepts.

Grouping. Following a pattern-based design, COMM already provides a rich degree of conceptual groupings and their axiomatization in a machine readable format. However, reusability can be improved by redistributing concepts among the six ontologies of COMM, *core*, *datatype*, *localization*, *media*, *visual*, and *textual* respectively. As an example, the concept RootSegmentRole, located in the COMM *core ontology*, is not used in any pattern definition and has therefore been relocated to the *localization ontology*.

Mapping. The main challenge of aligning COMM and the M3O concern the differences of the patterns used and how to relate them. Although some principles are shared between the ontologies, there are also major differences, e.g., the Digital Data Pattern of COMM and the Information Realization Pattern of the M3O.

Often COMM patterns have been replaced using a more generic pattern of the M3O. As an example, Figure 2 displays the adaptation of the COMM Digital Data Pattern through the M3O. For the alignment, we have decided that the functionality of the Digital Data Pattern, i.e., expressing data values, can be maintained by adopting the M3O Data Value Pattern instead. All related concepts have either been removed or mapped to the next matching M3O concept. As an example, the StructuredDataDescription concept has been removed as it held no further relevance in the context of the Data Value Pattern. The StructuredDataParameter concept on the other hand has been preserved as specialization of the M3O Annotation Pattern. To accommodate StructuredDataParameters with the M3O, we

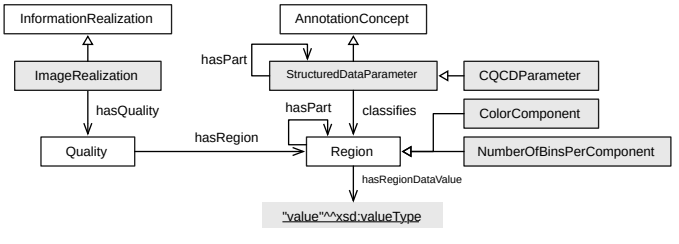


Fig. 2. Excerpt of COMM subsequently to the integration with the M3O. White boxes show the concepts of the M3O or DUL, whereas grey boxes represent concepts of COMM aligned to the M3O.