COMM: Designing a Well-Founded Multimedia Ontology for the Web

Richard Arndt[1], Raphaël Troncy[2], Steffen Staab[1], Lynda Hardman*[2], and Miroslav Vacura[3]

[1] ISWeb, University of Koblenz-Landau, Germany,
{rarndt|staab}@uni-koblenz.de
[2] CWI, Amsterdam, The Netherlands
{Raphael.Troncy|Lynda.Hardman}@cwi.nl
[3] University of Economics, Prague {vacuram@vse.cz}

Abstract. Semantic descriptions of non-textual media available on the web can be used to facilitate retrieval and presentation of media assets and documents containing them. While technologies for multimedia semantic descriptions already exist, there is as yet no formal description of a high quality multimedia ontology that is compatible with existing (semantic) web technologies. We explain the complexity of the problem using an annotation scenario. We then derive a number of requirements for specifying a formal multimedia ontology before we present the developed ontology, COMM, and evaluate it with respect to our requirements. We provide an API for generating multimedia annotations that conform to COMM.

1 Introduction

Multimedia objects on the Web are ubiquitous, whether found via web-wide search (e.g., Google or Yahoo! images[4]) or via dedicated sites (e.g., Flickr or YouTube[5]). These media objects are produced and consumed by professionals and amateurs alike. Unlike textual assets, whose content can be searched for using text strings, media search is dependent on processes that have either cumbersome requirements for feature comparison (e.g. color or texture) or rely on associated, more easily processable descriptions, selecting aspects of an image or video and expressing them as text, or as concepts from a predefined vocabulary. Individual annotation and tagging applications have not yet achieved a degree of interoperability that enables effective sharing of semantic metadata and that links the metadata to semantic data and ontologies found in the Semantic Web.

MPEG-7 [1, 2] is an international standard that specifies how to connect descriptions to parts of a media asset. The standard includes descriptors representing low-level media-specific features that can often be automatically extracted from media types. Unfortunately, MPEG-7 is not currently suitable for

* Lynda Hardman is also affiliated with the Technical University of Eindhoven.
[4] http://images.google.com/, http://images.search.yahoo.com/
[5] http://www.flickr.com/, http://www.youtube.com/

describing multimedia content on the Web, because *i)* its XML Schema-based nature prevents direct machine processing of semantic descriptions and its use of URNs is cumbersome for the Web; *ii)* it is not open to Web standards, which represent knowledge and make use of existing controlled vocabularies.

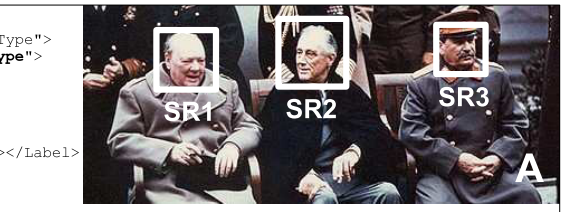
The Web provides an open environment where information can be shared and linked to. It has, however, no agreed-upon means of describing and connecting semantics with (parts of) multimedia assets and documents. While multimedia description frameworks, such as MPEG-7, already exist, no formal description of a multimedia ontology is compatible with existing (semantic) web technologies. Our contribution is thus to combine the advantages of the extensibility and scalability of web-based solutions with the accumulated experience of MPEG-7. Our approach advocates the use of formal semantics, grounded in a sound ontology development methodology, to describe the required multimedia semantics in terms of current semantic web languages. We develop COMM, a Core Ontology for MultiMedia.

In the next section, we illustrate the main problems when using MPEG-7 for describing multimedia resources on the web. In section 3, we review existing multimedia ontologies and show why the proposals made so far are inadequate for our purposes. Subsequently, we define the requirements that a multimedia ontology should meet (section 4) before we present COMM – an MPEG-7 based ontology, designed using sound design principles – and discuss our design decisions based on our requirements (section 5). In section 6, we demonstrate the use of the ontology with the scenario from section 2 and then conclude with some observations and future work.

2 Annotating Multimedia Documents on the Web

Let us imagine that an employee of an encyclopedia company wants to create a multimedia presentation of the Yalta Conference. For that purpose, s/he uses an MPEG-7 compliant authoring tool for detecting and labeling relevant multimedia objects automatically. On the web, the employee finds three different face recognition web services, each of them providing very good results for detecting Winston Churchill, Franklin D. Roosevelt and Josef Stalin respectively. Having these tools, the employee would like to run the face recognition web services on images and import the extraction results into the authoring tool in order to automatically generate links from the detected face regions to detailed textual information about Churchill, Roosevelt and Stalin. Fig. 1-A is an example of such an image; the bounding boxes are generated by the face recognition web services and linked to textual data by the authoring tool. This scenario, however, causes several problems with existing solutions:

Fragment identification. Particular regions of the image need to be localized (anchor value in [3]). However, the current web architecture does not provide a means for uniquely identifying sub-parts of multimedia assets, in the same way that the fragment identifier in the URI can refer to part of an HTML or XML document. Actually, for almost all other media types, the semantics of the




```
<Mpeg7>
 <Description xsi:type="ContentEntityType">
  <MultimediaContent xsi:type="ImageType">
   <Image id="IMG1">
    <SpatialDecomposition>

     <StillRegion id="SR1">
      <Semantic>
       <Label><Name> Roosevelt </Name></Label>
      </Semantic>
     </StillRegion>

     <StillRegion id="SR2">
      <TextAnnotation>     <!-- TextAnnotationType -->
       <KeywordAnnotation><Keyword> Churchill </Keyword></KeywordAnnotation>
      </TextAnnotation>
     </StillRegion>

     <StillRegion id="SR3">
      <Semantic>
       <Definition>   <!-- Also TextAnnotationType -->
        <StructuredAnnotation><Who><Name> Stalin </Name></Who></StructuredAnnotation>
       </Definition>
      </Semantic>
     </StillRegion>
     ...
```

B

Fig. 1. MPEG-7 annotation example (Image adapted from Wikipedia), `http://en.wikipedia.org/wiki/Yalta_Conference`

fragment identifier has not been defined or is not commonly accepted. Providing an agreed upon way to localize sub-parts of multimedia objects (e.g. sub-regions of images, temporal sequences of videos or tracking moving objects in space and in time) is fundamental[6] [4]. For images, one can use either MPEG-7 or SVG snippet code to define the bounding box coordinates of specific regions. For temporal location, one can use MPEG-7 code or the TemporalURI RFC[7]. MPEG-21 specifies a normative syntax to be used in URIs for addressing parts of any resource but whose media type is restricted to MPEG [5]. The MPEG-7 approach requires an indirection: an annotation is *about* a fragment of a XML document that *refers* to a multimedia document, whereas the MPEG-21 approach does not have this limitation.

Semantic annotation. MPEG-7 is a natural candidate for representing the extraction results of multimedia analysis software such as a face recognition web service. The language, standardized in 2001, specifies a rich vocabulary of multimedia descriptors, which can be represented in either XML or a binary format. While it is possible to specify very detailed annotations using these descriptors, it is not possible to guarantee that MPEG-7 metadata generated by different agents will be mutually understood due to the lack of formal semantics of this language [6, 7]. The XML code of Fig. 1-B illustrates the inherent interoperability problems of MPEG-7: several descriptors, semantically equivalent and representing the same information while using different syntax can coexist [8]. As our employee used three different face recognition web services, the extraction results of the regions `SR1`, `SR2` and `SR3` differ from each other even though they are all

[6] See also the related discussion in the W3C Multimedia Semantics XG `http://lists.w3.org/Archives/Public/public-xg-mmsem/2007Apr/0007.html`.

[7] `http://www.annodex.net/TR/URI_fragments.html`

syntactically correct. While the first service uses the MPEG-7 `SemanticType` for assigning the `<Label>` *Roosevelt* to still region `SR1`, the second one makes use of a `<KeywordAnnotation>` for attaching the keyword *Churchill* to still region `SR2`. Finally the third service uses a `<StructuredAnnotation>` (which can be used within the `SemanticType`) in order to label still region `SR3` with *Stalin*. Consequently, alternative ways for annotating the still regions render almost impossible the retrieval of the face recognition results within the authoring tool since the corresponding XPath query has to deal with these syntactic variations. As a result, the authoring tool will not link occurrences of Churchill in the image with, for example, his biography as it does not expect semantic labels of still regions behind the `<KeywordAnnotation>` element.

Web interoperability. Finally, our employee would like to link the multimedia presentation to historical information about the key figures of the Yalta Conference that is already available on the web. S/He has also found semantic metadata about the relationships between these figures that could improve the automatic generation of the multimedia presentation. However, s/he realizes that MPEG-7 cannot be combined with these concepts defined in domain-specific ontologies because of its closing to the web. As this example demonstrates, although MPEG-7 provides ways of associating semantics with (parts of) non-textual media assets, it is incompatible with (semantic) web technologies and has no formal description of the semantics encapsulated implicitly in the standard.

3 Related Work

In the field of semantic image understanding, using a multimedia ontology infrastructure is regarded to be the first step for closing the, so-called, semantic gap between low-level signal processing results and explicit semantic descriptions of the concepts depicted in images. Furthermore, multimedia ontologies have the potential to increase the interoperability of applications producing and consuming multimedia annotations. The application of multimedia reasoning techniques on top of semantic multimedia annotations is also a research topic which is currently investigated [9]. A number of drawbacks of MPEG-7 have been reported [10, 11]. As a solution, multimedia ontologies based on MPEG-7 have been proposed.

Hunter [6] provided the first attempt to model parts of MPEG-7 in RDFS, later integrated with the ABC model. Tsinaraki et al. [12] start from the core of this ontology and extend it to cover the full Multimedia Description Scheme (MDS) part of MPEG-7, in an OWL DL ontology. A complementary approach was explored by Isaac and Troncy [13], who proposed a core audio-visual ontology inspired by several terminologies such as MPEG-7, TV Anytime or ProgramGuideML. Garcia and Celma [14] produced the first complete MPEG-7 ontology, automatically generated using a generic mapping from XSD to OWL. Finally, Simou proposed an OWL DL Visual Descriptor Ontology[8] (VDO) based on the Visual part of MPEG-7 and used for image and video analysis.

[8] `http://image.ece.ntua.gr/~gstoil/VDO`

All these methods perform a one to one translation of MPEG-7 types into OWL concepts and properties. This translation does not, however, guarantee that the intended semantics of MPEG-7 is fully captured and formalized. On the contrary, the syntactic interoperability and conceptual ambiguity problems illustrated in section 2 remain.

4 Requirements for Designing a Multimedia Ontology

Requirements for designing a multimedia ontology have been gathered and reported in the literature, e.g. in [15]. Here, we compile these and use our scenario to present a list of requirements for a web-compliant multimedia ontology.

MPEG-7 compliance. MPEG-7 is an existing international standard, used both in the signal processing and the broadcasting communities. It contains a wealth of accumulated experience that needs to be included in a web-based ontology. In addition, existing annotations in MPEG-7 should be easily expressible in our ontology.

Semantic interoperability. Annotations are only re-usable when the captured semantics can be shared among multiple systems and applications. Obtaining similar results from reasoning processes about terms in different environments can only be guaranteed if the semantics is sufficiently explicitly described. A multimedia ontology has to ensure that the intended meaning of the captured semantics can be shared among different systems.

Syntactic interoperability. Systems are only able to share the semantics of annotations if there is a means of conveying this in some agreed-upon syntax. Given that the (semantic) web is an important repository of both media assets and annotations, a semantic description of the multimedia ontology should be expressible in a web language (e.g. OWL, RDF/XML or RDFa).

Separation of concerns. Clear separation of domain knowledge (i.e. knowledge about depicted entities, such as the person Winston Churchill) from knowledge that is related to the administrative management or the structure and the features of multimedia documents (e.g. Churchill's face is to the left of Roosevelt's face) is required. Reusability of multimedia annotations can only be achieved if the connection between both ontologies is clearly specified by the multimedia ontology.

Modularity. A complete multimedia ontology can be, as demonstrated by MPEG-7, very large. The design of a multimedia ontology should thus be made modular, to minimize the execution overhead when used for multimedia annotation. Modularity is also a good engineering principle.

Extensibility. While we intend to construct a comprehensive multimedia ontology, as ontology development methodologies demonstrate, this can never be complete. New concepts will always need to be added to the ontology. This requires a design that can always be extended, without changing the underlying model and assumptions and without affecting legacy annotations.

5 Adding Formal Semantics to MPEG-7

MPEG-7 specifies the connection between semantic annotations and parts of media assets. We take it as a base of knowledge that needs to be expressible in our ontology. Therefore, we re-engineer MPEG-7 according to the intended semantics of the written standard. We satisfy our semantic interoperability not by aligning our ontology to the XML Schema definition of MPEG-7, but by providing a formal semantics for MPEG-7. We use a methodology based on a foundational, or top level, ontology as a basis for designing COMM. This provides a domain independent vocabulary that explicitly includes formal definitions of foundational categories, such as processes or physical objects, and eases the linkage of domain-specific ontologies because of the definition of top level concepts. We briefly introduce our chosen foundational ontology in section 5.1, and then present our multimedia ontology, COMM, in sections 5.2 and 5.3. Finally, we discuss why our ontology satisfies all our stated requirements in section 5.4.

COMM is available at `http://multimedia.semanticweb.org/COMM/`.

5.1 DOLCE as Modeling Basis

Using the review in [16], we select the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [17] as a modeling basis. Our choice is influenced by two of the main design patterns: *Descriptions & Situations* (D&S) and *Ontology of Information Objects* (OIO) [18]. The former can be used to formalize contextual knowledge, while the latter, based on D&S, implements a semiotics model of communication theory. We consider that the annotation process is a *situation* (i.e. a reified context) that needs to be described.

5.2 Multimedia Patterns

The patterns for D&S and OIO need to be extended for representing MPEG-7 concepts since they are not sufficiently specialized to the domain of multimedia annotation. This section introduces these extended multimedia design patterns, while section 5.3 details two central concepts underlying these patterns: digital data and algorithms. In order to define design patterns, one has to identify repetitive structures and describe them at an abstract level. We have identified the two most important functionalities provided by MPEG-7 in the scenario presented in section 2: the *decomposition* of a media asset and the (semantic) *annotation* of its parts, which we include in our multimedia ontology.

Decomposition. MPEG-7 provides descriptors for spatial, temporal, spatio-temporal and media source decompositions of multimedia content into segments. A segment is the most general abstract concept in MPEG-7 and can refer to a region of an image, a piece of text, a temporal scene of a video or even to a moving object tracked during a period of time.

Annotation. MPEG-7 defines a very large collection of descriptors that can be used to annotate a segment. These descriptors can be low-level visual

features, audio features or more abstract concepts. They allow the annotation of the content of multimedia documents or the media asset itself.

In the following, we first introduce the notion of multimedia data and then present the patterns that formalize the decomposition of multimedia content into segments, or allow the annotation of these segments. The decomposition pattern handles the structure of a multimedia document, while the media annotation pattern, the content annotation pattern and the semantic annotation pattern are useful for annotating the media, the features and the semantic content of the multimedia document respectively.

Multimedia Data. This encapsulates the MPEG-7 notion of multimedia content and is a subconcept of DigitalData[9] (introduced in more detail in section 5.3). MultimediaData is an abstract concept that has to be further specialized for concrete multimedia content types (e.g. ImageData corresponds to the pixel matrix of an image). According to the OIO pattern, MultimediaData is realized by some physical Media (e.g. an Image). This concept is needed for annotating the physical realization of multimedia content (see section 5.3).

Decomposition Pattern. Following the D&S pattern, we consider that a decomposition of a MultimediaData entity is a Situation (a SegmentDecomposition) that satisfies a Description, such as a SegmentationAlgorithm or a Method (e.g. a user drawing a bounding box around a depicted face), which has been applied to perform the decomposition, see Fig. 2-B. Of particular importance are the Roles that are defined by a SegmentationAlgorithm or a Method. OutputSegmentRoles express that some MultimediaData entities are segments of a MultimediaData entity that plays the role of an input segment (InputSegmentRole). These data entities have as setting a SegmentDecomposition situation that satisfies the roles of the applied SegmentationAlgorithm or Method. OutputSegmentRoles as well as SegmentDecompositions are then specialized according to the segment and decomposition hierarchies of MPEG-7 ([1], part 5, section 11).

The decomposition pattern also reflects the need for localizing segments within the input segment of a decomposition as each OutputSegmentRole requires a MaskRole. Such a role has to be played by one or more DigitalData entities which express one LocalizationDescriptor. An example of such a descriptor is an ontological representation of the MPEG-7 `RegionLocatorType`[10] for localizing regions in an image (see Fig. 2-C, details in section 5.3). Hence, the MaskRole concept corresponds to the notion of a mask in MPEG-7.

Content Annotation Pattern. This formalizes the attachment of metadata (i.e. annotations) to MultimediaData (Fig. 2-D). Using the D&S pattern, Annotations also become Situations that represent the state of affairs of all related DigitalData (metadata and annotated MultimediaData). DigitalData entities represent the attached metadata by playing an AnnotationRole. These Roles are defined

[9] Sans serif font indicates ontology concepts.
[10] Type writer font indicates MPEG-7 language descriptors.

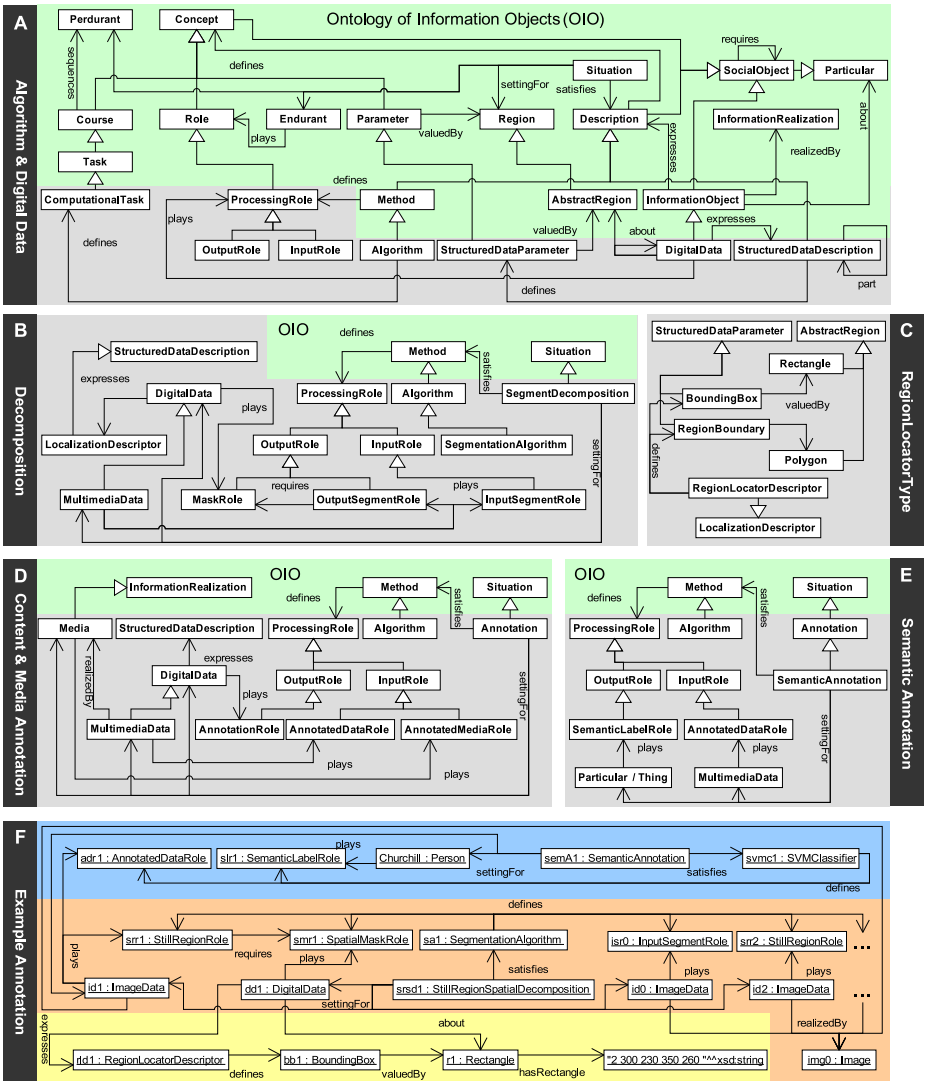


Fig. 2. COMM: Design patterns in UML notation: Basic design patterns (A), multimedia patterns (B, D, E), modeling example (C) and example annotation graph (F).

by Methods or Algorithms. The former are used to express manual (or semi-automatic) Annotation while the latter serve as an explanation for the attachment of automatically computed features, such as the dominant colors of a still region. It is mandatory that the MultimediaData entity being annotated plays an AnnotatedDataRole.

The actual metadata that is carried by a DigitalData entity depends on the StructuredDataDescription that is expressed by it. These descriptions are for-

malized using the digital data pattern (see section 5.3). Applying the content annotation pattern for formalizing a specific annotation, e.g. a DominantColorAnnotation which corresponds to the connection of a MPEG-7 `DominantColorType` with a segment, requires only the specialization of the concept Annotation, e.g. DominantColorAnnotation. This concept is defined by being a setting for a DigitalData entity that expresses one DominantColorDescriptor (a subconcept of StructuredDataDescription which corresponds to the `DominantColorType`).

Media Annotation Pattern. This forms the basis for describing the physical instances of multimedia content (Fig. 2-D). It differs from the content annotation pattern in only one respect: it is the Media that is being annotated and therefore plays an AnnotatedMediaRole.

One can thus represent that the content of Fig. 1-A is realized by a JPEG Image with a size of 462848 byte, using the MPEG-7 `MediaFormatType`. Using the media annotation pattern, the metadata is attached by connecting a DigitalData entity with the Image. The DigitalData plays an AnnotationRole while the Image plays an AnnotatedMediaRole. An ontological representation of the `MediaFormatType`, namely an instance of the StructuredDataDescription subconcept MediaFormatDescriptor, is expressed by the DigitalData entity. The tuple formed with the scalar "462848" and the string "JPEG" is the value of the two instances of the concepts FileSize and FileFormat respectively. Both concepts are subconcepts of StructuredDataParameter (Fig. 2-C).

Semantic Annotation Pattern. Even though MPEG-7 provides some general concepts (see [1], part 5, section 12) that can be used to describe the perceivable content of a multimedia segment, independent development of domain-specific ontologies is more appropriate for describing possible interpretations of multimedia — it is useful to create an ontology specific to multimedia, it is not useful to try to model the real world within this. An ontology-based multimedia annotation framework should rely on domain-specific ontologies for the representation of the real world entities that might be depicted in multimedia content. Consequently, this pattern specializes the content annotation pattern to allow the connection of multimedia descriptions with domain descriptions provided by independent world ontologies (Fig. 2-E).

An OWL Thing or a DOLCE Particular (belonging to a domain-specific ontology) that is depicted by some multimedia content is not directly connected to it but rather through the way the annotation is obtained. Actually, a manual annotation Method or its subconcept Algorithm, such as a classification Algorithm, has to be applied to determine this connection. It is embodied through a SemanticAnnotation that satisfies the applied Method. This Description specifies that the annotated MultimediaData has to play an AnnotatedDataRole and the depicted Thing / Particular has to play a SemanticLabelRole. The pattern also allows the integration of features which might be evaluated in the context of a classification Algorithm. In that case, DigitalData entities that represent these features would play an InputRole.

5.3 Basic Patterns

Specializing the D&S and OIO patterns for defining multimedia design patterns is enabled through the definition of basic design patterns, which formalize the notion of digital data and algorithm.

Digital Data Pattern. Within the domain of multimedia annotation, the notion of digital data is central — both the multimedia content being annotated and the annotations themselves are expressed as digital data. We consider DigitalData entities of arbitrary size to be InformationObjects, which are used for communication between machines. The OIO design pattern states that Descriptions are expressed by InformationObjects, which have to be about facts (represented by Particulars). These facts are settings for Situations that have to satisfy the Descriptions that are expressed by InformationObjects. This chain of constraints allows the modeling of complex data structures to store digital information. Our approach is as follows (see Fig. 2-A): DigitalData entities express Descriptions, namely StructuredDataDescriptions, which define meaningful labels for the information contained by DigitalData. This information is represented by numerical entities such as scalars, matrices, strings, rectangles or polygons. In DOLCE terms, these entities are AbstractRegions. In the context of a Description, these Regions are described by Parameters. StructuredDataDescriptions thus define StructuredDataParameters, for which AbstractRegions carried by DigitalData entities assign values.

The digital data pattern can be used to formalize complex MPEG-7 low-level descriptors. Fig. 2-C shows the application of this pattern by formalizing the MPEG-7 `RegionLocatorType`, which mainly consists of two elements: a `Box` and a `Polygon`. The concept RegionLocatorDescriptor corresponds to the `RegionLocatorType`. The element `Box` is represented by the StructuredDataParameter subconcept BoundingBox while the element `Polygon` is represented by the RegionBoundary concept.

The MPEG-7 code example given in Fig. 1 highlights that the formalization of data structures, so far, is not sufficient — complex MPEG-7 types can include nested types that again have to be represented by StructuredDataDescriptions. In our example, the MPEG-7 `SemanticType` contains the element `Definition` which is of complex type `TextAnnotationType`. The digital data pattern covers such cases by allowing a DigitalData instance dd1 to be about a DigitalData instance dd2 which expresses a StructuredDataDescription that corresponds to a nested type (see Fig. 2-A). In this case the StructuredDataDescription of instance dd2 would be a part of the one expressed by dd1.

Algorithm Pattern. The production of multimedia annotation can involve the execution of Algorithms or the application of computer assisted Methods which are used to produce or manipulate DigitalData. The automatic recognition of a face in an image region is an example of the former, while manual annotation of the characters is an example of the latter.

We consider Algorithms to be Methods that are applied to solve a computational problem (see Fig. 2-A). The associated (DOLCE) Situations represent

the work that is being done by Algorithms. Such a Situation encompasses DigitalData[11] involved in the computation, Regions that represent the values of Parameters of an Algorithm, and Perdurants[12] that act as ComputationalTasks (i.e. the processing steps of an Algorithm). An Algorithm defines Roles which are played by DigitalData. These Roles encode the meaning of data. In order to solve a problem, an Algorithm has to process input data and return some output data. Thus, every Algorithm defines at least one InputRole and one OutputRole which both have to be played by DigitalData.

5.4 Comparison with Requirements

We discuss now whether the requirements stated in section 4 are satisfied with our proposed modeling of the multimedia ontology.

The ontology is **MPEG-7 compliant** since the patterns have been designed with the aim of translating the standard into DOLCE. It covers the most important part of MPEG-7 that is commonly used for describing the structure and the content of multimedia documents. Our current investigation shows that parts of MPEG-7 which have not yet been considered (e.g. navigation & access) can be formalized analogously to the other descriptors through the definition of further patterns. The technical realization of the basic MPEG-7 data types (e.g. matrices and vectors) is not within the scope of the multimedia ontology. They are represented as ontological concepts, because the about relationship which connects DigitalData with numerical entities is only defined between concepts. Thus, the definition of OWL data type properties is required to connect instances of data type concepts (subconcepts of the DOLCE AbstractRegion) with the actual numeric information (e.g. xsd:string). Currently, simple string representation formats are used for serializing data type concepts (e.g. Rectangle) that are currently not covered by W3C standards. Future work includes the integration of the extended data types of OWL 1.1.

Syntactic and semantic interoperability of our multimedia ontology is achieved by an OWL DL formalization[13]. Similar to DOLCE, we provide a rich axiomatization of each pattern using first order logic. Our ontology can be linked to any web-based domain-specific ontology through the semantic annotation pattern.

A clear **separation of concerns** is ensured through the use of the multimedia patterns: the decomposition pattern for handling the structure and the annotation pattern for dealing with the metadata.

These patterns form the core of the **modular** architecture of the multimedia ontology. We follow the various MPEG-7 parts and organize the multimedia ontology into modules which cover *i)* the descriptors related to a specific media type (e.g. visual, audio or text) and *ii)* the descriptors that are generic to a

[11] DigitalData entities are DOLCE Endurants, i.e. entities which exist in time and space.
[12] Events, processes or phenomena are examples of Perdurants. Endurants participate in Perdurants.
[13] Examples of the axiomatization are available on the COMM website.

particular media (e.g. media descriptors). We also design a separate module for data types in order to abstract from their technical realization.

Through the use of multimedia design patterns, our ontology is also **extensible**, allowing the inclusion of further media types and descriptors (e.g. new low-level features) using the same patterns. As our patterns are grounded in the D&S pattern, it is straightforward to include further contextual knowledge (e.g. about provenance) by adding Roles or Parameters. Such extensions will not change the patterns, so that legacy annotations will remain valid.

6 Expressing the Scenario in COMM

The interoperability problem with which our employee was faced in section 2 can be solved by employing the COMM ontology for representing the metadata of all relevant multimedia objects and the presentation itself throughout the whole creation workflow. The employee is shielded from details of the multimedia ontology by embedding it in authoring tools and feature analysis web services.

The application of the Winston Churchill face recognizer results in an annotation RDF graph that is depicted in Fig. 2-F (visualized by an UML object diagram[14]). The decomposition of Fig. 1-A, whose content is represented by id0, into one still region (the bounding box of Churchill's face) is represented by the large middle part of the UML diagram. The segment is represented by the ImageData instance id1 which plays the StillRegionRole srr1. It is located by the DigitalData instance dd1 which expresses the RegionLocatorDescriptor rld1 (lower part of the diagram). Due to the semantic annotation pattern, the face recognizer can annotate the still region by connecting it with the instance Churchill of a domain ontology that contains historic Persons (upper part of Fig. 2-F).

Running the two remaining face recognizers for Roosevelt and Stalin will extend the decomposition further by two still regions, i.e. the ImageData instances id2 and id3 as well as the corresponding StillRegionRoles, SpatialMaskRoles and DigitalData instances expressing two more RegionLocatorDescriptors (indicated at the right border of Fig. 2-F). The domain ontologies which provide the instances Roosevelt and Stalin for annotating id2 and id3 with the semantic annotation pattern do not have to be identical to the one that contains Churchill. If several domain ontologies are used, the employee can use the OWL sameAs and equivalentClass constructs to align the three face recognition results to the domain ontology that is best suited for enhancing the automatic generation of the multimedia presentation.

In order to ease the creation of multimedia annotations with our ontology, we have developed a Java API[15] which provides an MPEG-7 class interface for the construction of meta-data at runtime. Annotations which are generated in memory can be exported to Java based RDF triple stores such as Sesame. For that purpose, the API translates the objects of the MPEG-7 classes into instances of the COMM concepts. The API also facilitates the implementation

[14] The scheme used in Fig. 2-F is instance:Concept, the usual UML notation.

[15] The Java API is available at `http://multimedia.semanticweb.org/COMM/api/`.

of multimedia retrieval tools as it is capable of loading RDF annotation graphs (e.g. the complete annotation of an image including the annotation of arbitrary regions) from a store and converting them back to the MPEG-7 class interface. Using this API, the face recognition web service will automatically create the annotation which is depicted in Fig. 2-F by executing the following code:

```
Image img0 = new Image();
StillRegion isr0 = new StillRegion();
img0.setImage(isr0);
StillRegionSpatialDecomposition srsd1 = new StillRegionSpatialDecomposition();
isr0.addSpatialDecomposition(srsd1);
srsd1.setDescription(new SegmentationAlgorithm());
StillRegion srr1 = new StillRegion();
srsd1.addStillRegion(srr1);
SpatialMask smr1 = new SpatialMask();
srr1.setSpatialMask(smr1);
RegionLocatorDescriptor rld1 = new RegionLocatorDescriptor();
smr1.addSubRegion(rld1);
rld1.setBox(new Rectangle(300, 230, 50, 30));
Semantic s1 = new Semantic();
s1.addLabel("http://en.wikipedia.org/wiki/Winston_Churchill");
s1.setDescription(new SVMClassifier());
srr1.addSemantic(s1);
```

7 Conclusion and Future Work

We have developed COMM, an MPEG-7 based multimedia ontology, well-founded, composed of multimedia patterns. This satisfies the requirements, as they are described by the multimedia community itself, for a multimedia ontology framework. The ontology is completely formalized in OWL DL and a stable version is available with its API at: `http://multimedia.semanticweb.org/COMM/`.

The ontology already covers the main parts of the standard, and we are confident that the remaining parts can be covered by following our method for extracting more design patterns. Our modeling approach confirms that the ontology offers even more possibilities for multimedia annotation than MPEG-7 since it is interoperable with existing web ontologies. The explicit representation of algorithms in the multimedia patterns describes the multimedia analysis steps, something that is not possible in MPEG-7. The need for providing this kind of annotation is demonstrated in the use cases of the W3C Multimedia Semantics Incubator Group[16]. The intensive use of the D&S reification mechanism causes that RDF annotation graphs, which are generated according to our ontology, are quite large compared to the ones of more straightforwardly designed multimedia ontologies. However, the situated modeling allows COMM to represent very general annotations. Future work will improve our evaluation of the ontology, its scalability and its adequacy in the implementation of tools that use it for multimedia annotation, analysis and reasoning in large scale applications.

[16] `http://www.w3.org/2005/Incubator/mmsem/XGR-interoperability/`

Acknowledgments

The research leading to this paper was partially supported by the European Commission under contract FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content – K-Space and under contract FP6-026978, X-Media Integrated Project.

References

[1] MPEG-7: Multimedia Content Description Interface. ISO/IEC 15938 (2001)
[2] Nack, F., Lindsay, A.T.: Everything you wanted to know about MPEG-7 (Parts I & II). IEEE Multimedia **6**(3-4) (1999)
[3] Halasz, F., Schwartz, M.: The Dexter Hypertext Reference Model. Communications of the ACM **37**(2) (1994) 30–39
[4] Geurts, J., Ossenbruggen, J.v., Hardman, L.: Requirements for practical multimedia annotation. In: Workshop on Multimedia and the Semantic Web. (2005)
[5] MPEG-21: Part 17: Fragment Identification of MPEG Resources. Standard No. ISO/IEC 21000-17 (2006)
[6] Hunter, J.: Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In: 1^{st} Int. Semantic Web Working Symposium. (2001) 261–281
[7] Troncy, R.: Integrating Structure and Semantics into Audio-visual Documents. In: 2^{nd} Int. Semantic Web Conference. (2003) 566–581
[8] Troncy, R., Bailer, W., Hausenblas, M., Hofmair, P., Schlatte, R.: Enabling Multimedia Metadata Interoperability by Defining Formal Semantics of MPEG-7 Profiles. In: 1^{st} Int. Conf. on Semantics And digital Media Technology. (2006) 41–55
[9] Neumann, B., Möller, R.: On Scene Interpretation with Description Logics. In: Cognitive Vision Systems. Springer (2006) 247–275
[10] Ossenbruggen, J.v., Nack, F., Hardman, L.: That Obscure Object of Desire: Multimedia Metadata on the Web (Part I). IEEE Multimedia **11**(4) (2004)
[11] Nack, F., Ossenbruggen, J.v., Hardman, L.: That Obscure Object of Desire: Multimedia Metadata on the Web (Part II). IEEE Multimedia **12**(1) (2005)
[12] Tsinaraki, C., Polydoros, P., Moumoutzis, N., Christodoulakis, S.: Integration of OWL ontologies in MPEG-7 and TV-Anytime compliant Semantic Indexing. In: 16^{th} Int. Conference on Advanced Information Systemes Engineering. (2004)
[13] Isaac, A., Troncy, R.: Designing and Using an Audio-Visual Description Core Ontology. In: Workshop on Core Ontologies in Ontology Engineering. (2004)
[14] Garcia, R., Celma, O.: Semantic Integration and Retrieval of Multimedia Metadata. In: 5^{th} Int. Workshop on Knowledge Markup and Semantic Annotation. (2005)
[15] Hunter, J., Armstrong, L.: A Comparison of Schemas for Video Metadata Representation. In: 8^{th} Int. World Wide Web Conference. (1999) 1431–1451
[16] Oberle, D., Lamparter, S., Grimm, S., Vrandecic, D., Staab, S., Gangemi, A.: Towards Ontologies for Formalizing Modularization and Communication in Large Software Systems. Journal of Applied Ontology **1**(2) (2006) 163–202
[17] Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., Schneider, L.: The WonderWeb Library of Foundational Ontologies (WFOL). Technical report, WonderWeb Deliverable 17 (2002)
[18] Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C., Ferrario, R., Gangemi, A., Guarino, N.: Social Roles and their Descriptions. In: 9^{th} Int. Conference on Principles of Knowledge Representation and Reasoning. (2004) 266–277